R E P O R T   R E S U M E S

ED 003 442 24
VARIABLES RELATED TO OUTCOMES OF TEST INTERPRETATION.
BY- OHLSEN, MERLE M. AND OTHERS
ILLINOIS UNIV., URBANA, COLL. OF EDUCATION
REPORT NUMBER CRP-S-024 PUB DATE 64
EDRS PRICE MF-$0.18 HC-$3.88 97P.

DESCRIPTORS- *STANDARDIZED TESTS, *COUNSELING, *TEST INTERPRETATION, SELF CONCEPT, SELF EVALUATION, TEACHER EVALUATION, *ELEMENTARY SCHOOL STUDENTS, *HIGH SCHOOL STUDENTS, URBANA, ILLINOIS

THIS THREE-PART STUDY WAS CONCERNED WITH THE MANNER IN WHICH CERTAIN VARIABLES SEEMED TO BE RELATED TO THE OUTCOMES OF TEST INTERPRETATION AND TO STUDENTS' PERCEPTION OF THEMSELVES AND THEIR TEACHERS' PERCEPTION OF THEM. PART ONE WAS CONCERNED WITH THE OUTCOMES OF TEST INTERPRETATION AS RELATED TO SUCH FACTORS AS (A) COUNSELOR WHO INTERPRETED THE TEST SCORES, (B) SCHOOL MARKS, (C) SELF-ESTIMATES, (D) ACCURACY OF SELF-ESTIMATES, (E) FATHER'S OCCUPATIONAL LEVEL, (F) FATHER'S EDUCATIONAL LEVEL, (G) MOTHER'S EDUCATIONAL LEVEL AND OTHERS. PART TWO WAS CONCERNED WITH FACTORS ASSOCIATED WITH TEACHER OVER- AND UNDER-ESTIMATION OF STUDENTS' INTELLECTUAL ABILITY. PART THREE INVESTIGATED THE RELATIONSHIP BETWEEN TEACHERS' ESTIMATES OF INTELLECTUAL ABILITY AND THE STUDENTS' SELF-ESTIMATES. DATA WERE USED FROM A PREVIOUS COOPERATIVE RESEARCH PROJECT (NUMBER 1344) ON 703 STUDENTS IN GRADES 5, 7, 9, AND 11 FROM 5 RURAL SCHOOL DISTRICTS. FINDINGS COULD BE USED BY COUNSELORS TO BETTER UNDERSTAND STUDENTS' BEHAVIOR AND TO HELP COUNSELORS DEAL WITH STUDENTS' REACTION TO WHAT THEY LEARN ABOUT THEMSELVES. IN PARTICULAR, THESE RESULTS SUGGEST THAT COUNSELORS SPEND MORE TIME WITH THE UNDERESTIMATORS. (JM)

ED003442

UNIVERSITY OF ILLINOIS

VARIABLES RELATED TO OUTCOMES OF TEST INTERPRETATION

Cooperative Research Project

S-024

No. ~~S-043-64~~

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

Merle M. Ohlsen
Richard E. Pearson
Paul A. Wurm

1964

COLLEGE OF EDUCATION
UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

VARIABLES RELATED TO OUTCOMES OF TEST INTERPRETATION

by

Merle M. Ohlsen
Richard E. Pearson
Paul A. Wurm

College of Education
University of Illinois
Urbana, Illinois

024
Cooperative Research Project No. S-~~043-64~~

The research reported herein was supported by the Cooperative Research Program, Office of Education, United States Department of Health, Education, and Welfare.

1964

TABLE OF CONTENTS

SUMMARY
Merle M. Ohlsen 1

VARIABLES RELATED TO OUTCOMES OF TEST INTERPRETATION
Paul A. Wurm and Merle M. Ohlsen 9

FACTORS ASSOCIATED WITH TEACHER OVER- OR UNDERESTIMATION OF STUDENT INTELLECTUAL ABILITY
Richard E. Pearson and Merle M. Ohlsen 61

THE RELATIONSHIP BETWEEN TEACHERS' ESTIMATES OF STUDENTS' INTELLECTUAL ABILITY AND THE STUDENTS' SELF-ESTIMATES
Richard E. Pearson and Paul A. Wurm 80

LIST OF TABLES

Analysis of Variance of Congruency Scores of Students Seen by Different Counselors 21

Product-Moment Correlations Among Self-Estimates Test Scores, and School Marks 23

Analysis of Variance of Accuracy Scores of Students From Six Father's Occupational Groups 25

Analysis of Variance of Perceived Accuracy Scores of Students From Six Father's Occupational Groups 25

Analysis of Variance of Satisfaction Scores of Students From Six Father's Occupational Groups 26

Analysis of Variance of Discrepancy Scores of Students From Six Father's Occupational Groups 27

Analysis of Variance of Congruency Scores of Students From Six Father's Occupational Groups 28

Relationship Between Father's Occupational Group and Motivation 30

t-Tests for Difference in Intelligence Scores of the High- and Low-Accuracy Criterion Groups 31

t-Tests For Difference In Perceived Accuracy Scores of the High- and Low-Accuracy Criterion Groups 32

t-Tests for Difference in Satisfaction Scores of the High- and Low-Accuracy Criterion Groups 33

t-Tests for Difference in Discrepancy Scores of the High- and Low-Accuracy Criterion Groups 33

t-Tests for Difference in Congruency Scores of the High- and Low-Accuracy Criterion Groups 33

Relationship Between Accuracy Rank and Father's Occupational Group 36

Relationship Between Accuracy Rank and Father's Educational Level 37

Relationship Between Accuracy Rank and Mother's Educational Level 37

Relationship Between Accuracy Rank and Motivation for Test Interpretation 37

t-Tests for Difference in Intelligence Scores of the High- and Low-Congruency Criterion Groups 40

t-Tests for Difference in Perceived Accuracy Scores of the High- and Low-Congruency Criterion Groups 40

LIST OF TABLES (continued)

t-Tests for Difference in Satisfaction Scores of the High- and Low-Congruency Criterion Groups 40

t-Tests for Difference in Discrepancy Scores of the High- and Low-Congruency Criterion Groups 41

Relationship Between Congruency Rank and Father's Occupational Group 43

Relationship Between Congruency Rank and Father's Educational Level 43

Relationship Between Congruency Rank and Mother's Educational Level 43

Relationship Between Congruency Rank and Motivation for Test Interpretation 44

Analysis of Variance of Intelligence Scores of Students From Three Estimation Ranks 45

Analysis of Variance of Accuracy Scores of Students From Three Estimation Ranks 46

Analysis of Variance of Perceived Accuracy Scores of Students From Three Estimation Ranks 48

Analysis of Variance of Satisfaction Scores of Students From Three Estimation Ranks 49

Analysis of Variance of Discrepancy Scores of Students From Three Estimation Ranks 50

Analysis of Variance of Congruency Scores of Students From Three Estimation Ranks 51

Relationship Between Estimation Rank and Father's Occupational Group 53

Relationship Between Estimation Rank and Father's Educational Level 53

Relationship Between Estimation Rank and Mother's Educational Level 53

Relationship Between Estimation Rank and Motivation for Test Interpretation 54

Relationship Between Teachers' Ratings of Students in Four Descriptive Areas, and Discrepancy Scores 67

Intercorrelation Matrix For the Five Areas In Which Teacher Ratings Were Obtained 68

t-Test of Mean Difference Between the Discrepancy Scores of Students Rated High and Those Rated Low in The Area of Personal Appearance 70

LIST OF TABLES (continued)

t-Test of Mean Difference Between the Discrepancy Scores of Students Rated High and Those Rated Low In the Area of Classroom Behavior 71

t-Test of Mean Differences Between the Discrepancy Scores of Students Rated High and Those Rated Low In the Area of Social and Economic Background 71

t-Test of Mean Difference Between the Discrepancy Scores of Students Rated High and Those Rated Low in the Area of Academic Achievement 72

Product-Moment Correlation Between Teachers' Estimate of Students' Intellectual Ability and The Students' Henmon-Nelson Performance 82

Significance of Differences Between Product-Moment Correlation of Henmon-Nelson Results and Teachers' Estimates of Students' Intellectual Ability, and Henmon-Nelson Results and Students' Self-Estimates 84

t-Test of Mean Difference Between Absolute Discrepancy of Teachers' Estimates of Students' Intellectual Ability and That of Students' Self-Estimates 85

Significance of Difference Between Product-Moment Correlation; of Self-Estimate and Henmon-Nelson Results and Self-Estimates and Teachers' Estimates 86

Relationship Between Classification of Students' Self-Estimates and That of Teachers' Estimates 88

SUMMARY

by Merle M. Ohlsen

Title: Variables Related to Outcomes of Test Interpretation

Investigators: Merle M. Ohlsen, Richard E. Pearson, and Paul A. Wurm

Institution: College of Education
University of Illinois
Urbana, Illinois

Project Number: S-043-64

Duration: February 1, 1964 to September 30, 1964

Background:

This study was concerned with the manner in which certain variables seemed to be related to the outcomes of test interpretation and to students' perception of themselves and their teachers' perception of them. It called for further analysis of test data and students' self-estimates collected for Lister and Ohlsen's[1] study and for analysis of teachers' descriptions of students and of data provided by parents which were collected for Lister and Ohlsen's study but not analyzed at that time. It also involved some new data on students' grades which were collected for this study.

Most of the studies concerned with outcomes of test interpretation have used either high school or college students as subjects, and rarely has increased congruency between students' self-estimates and test-estimates been obtained by investigators who used subjects below grade eleven. Lister and Ohlsen's study is one of the few exceptions. They found that (1) at all four grade levels (five, seven, nine, and eleven) students improved the accuracy of their self estimates (improved congruency between self-estimates and test estimates) for all three types of tests following test interpretation, and that students maintained a significantly more accurate self-estimate over the two month period following the test interpretation; (2) most students wanted to have their tests interpreted; (3) only limited support was found for the hypothesis that motivation for test interpretation increased accuracy of self-estimates following test interpretation; and (4) students' satisfaction with self-estimates and perceived accuracy of self-estimates increased over the four self-ratings (pre-testing, post-testing, post-test interpretations and follow-up) suggesting increased acceptance of test information. Perhaps four points should be noted about their study: (1) the counselors were selected with care to insure that they were well qualified of their duties; (2) they were very carefully prepared for their duties and tape recordings of their interviews were checked periodically to determine whether they followed the prescribed procedures; (3) students participated in the test interpretation process; and (4) counselors tried to detect how students felt about their test results and to help them discuss their reactions to what they learned about themselves.

---

[1] James L. Lister and Merle M. Ohlsen, *The Effects of Orientation to Testing on Motivation for and Outcomes of Test Interpretation*, Cooperative Research Project No. 1344, College of Education, University of Illinois, Urbana, 1962.

Thus increased self understanding can be achieved, and it is essential in order for students to determine what they have a right to expect from themselves, to understand and to cope with learning problems, and to define appropriate educational and vocational goals while they are enrolled in secondary schools, and after they leave school to revise these goals or to define new goals is the light of increased knowledge of self and changing economic conditions. It also is important that teachers understand their students in order to select appropriate materials, to plan meaningful learning experiences, and to help students diagnose and cope with learning problems. Furthermore, it is logical to assume that teachers' perceptions of their students effects students perception of themselves. Hence, it was essential that these relationships be investigated.

Objectives

Since this project has three independent parts the objectives for each are listed separately.

The objectives of part one were concerned with outcomes of test interpretation: (1) to investigate the extent to which outcomes of test interpretations vary with the counselor who interpreted the test scores; (2) to explore the relationships among school marks, self-estimates, and test scores; (3) to examine the relationship between father's occupational level and (a) accuracy of self-estimates, (b) perceived accuracy of self-estimates, (c) discrepancies between actual and ideal self-estimates, (d) outcomes of test interpretations, and (e) motivation for test interpretation; (4) to compare those students with most accurate self-estimates and those with least accurate self-estimates with reference to (a) intelligence, (b) perceived accuracy of self-estimates, (c) satisfaction with self-estimates, (d) discrepancies between actual and ideal self-estimates, (e) outcomes of test interpretation, (f) father's occupational level, (g) father's educational level, (h) mother's educational level, and (i) motivation for test interpretation; (5) to compare those students with greatest gains in accuracy (increased congruency) over test interpretation and those with least gains in accuracy on all the variables listed for (4) except (e); and (6) to compare those students with most accurate self-estimates with both over-estimators and underestimators on all the variables listed under (4).

Part two was concerned with factors associated with teacher over- and underestimation of students' intellectual ability.

Part three investigated the relationship between teachers' estimates of students' intellectual ability and the students' self-estimates. First the investigators determined whether chance could account for the observed product-moment correlation coefficient between teachers' estimates of students' intellectual ability and test estimates of it and for the observed difference between two pairs of correlation coefficients: (1) teachers' estimates and test estimates versus students self-estimates and test estimates and (2) students' self-estimates and test estimates versus students' self-estimates and teachers' estimates. Finally, part three was concerned with the relationship between students classified as over-estimated, accurately estimated, and under-estimated on the basis of students' self estimates and teachers' estimates.

Procedure

The students who participated in this study were enrolled in five rural school districts located within twenty miles of the University of Illinois. Within these communities live a good cross-section of all occupational groups, including slightly more professional people than in the typical Illinois towns of similar size. Of the 760 who participated in the original study, it was possible to collect complete data necessary for the first paper on only 703 students: 190 fifth graders, 176 seventh graders, 204 ninth graders, and 133 eleventh graders.

Students enrolled in one of the school districts which was used as a control in the original study were not included for the second and third parts of this project because the additional data were not collected there during the original study. Hence, only 562 students and 48 teachers participated in these phases of the project.

From the objectives it is evident that this was empirical study. Test estimates of mental ability were obtained on Henmon-Nelson Test of Mental Ability for all four grade levels. Test estimates of school achievement were obtained from Iowa Tests of Basic Skills, Form 1, for grades five and seven and from Iowa Tests of Educational Development, Form X3-5 (test 1, Background in the Social Sciences; Test 2, General Background in the Natural Sciences; Test 3, Correctness and Appropriateness of Expression; Test 4, Ability to do Quantitative Thinking; and Test 8, General Vocabulary) for grades nine and eleven. All tests were administered for the original study in late October and early November. At the time of testing, and prior to returning test results to the schools, Pearson obtained teachers' descriptions of their students' personal appearance, classroom conduct, social and economic background, intellectual ability, and academic achievement level.

The Self-Rating Scale used to obtain students' self-estimates in the original study was designed in two forms: one for grades five and seven and another for grades nine and eleven. It attempted to describe, in language appropriate for the grade level, the specific areas for which students were required to make self-estimates. For each score each student was required to make four self-evaluations: (1) describe himself as he thought he performed on that test or sub-test; (2) report his satisfaction with that description of himself; (3) indicate how he wished he were with respect to that area--his ideal self; and (4) indicate how accurately he felt that he had described himself with reference to that specific characteristic. Finally, these data were collected four times: pre-testing, post-testing, post-test interpretation, and after a two-month follow up period. Post-testing (or pre-test interpretation) and follow-up ratings were used in this project. However, pre-testing ratings did not differ significantly from post-testing ratings.

At the time the original data were collected parents completed a questionnaire which provided the essential data for classifying parents with reference to father's occupational level, father's educational level, and mother's educational level. Grades were obtained for the appropriate grading period especially for this study.

Four basic statistical procedures were used: (1) Analysis of Variance was used where the investigators wanted to check on possible significant interaction variance and where Kramer's[2] extension of Duncan's new multiple range could be applied to make multiple comparisons among group means where significant F ratios were obtained--e.g., to determine whether the counselor variable effected outcomes of test interpretations or to determine how other variables were effected when students were classified on the basis of father's occupational level; (2) t-tests were carried out independently for each grade level and for the total sample for such a purpose as to compare other variables for students who were classified in the highest and lowest fourth of their group on the basis of their accuracy scores; (3) Chi Square was used to test hypothesis involving data which were assigned to categories or cells, and hence were not continuous--e.g., for comparing levels of motivation for test interpretation for students whose parents were classified on the basis of fathers' occupational levels or for comparing fathers' occupational levels, fathers' educational levels, and mothers educational levels for students with accuracy scores in highest and lowest fourth of their groups; and (4) product-moment correlation coefficient were compared to study relationships between various combinations of variables listed earlier under the objectives. Where the stated hypothesis called for a comparison of correlation coefficients, procedures described by Walker and Lev[3] were used to determine whether chance would account for the observed differences in correlation coefficients.

Results and Conclusions

Although the results from this project are reported in three independent papers, each significant finding is summarized here and discussed, including its implications for school practices. Where a finding reported later in this summary of the research either contradicts or supports a finding discussed earlier reference is made to the earlier finding.

---

[2] C. Y. Kramer, "Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications, Biometrics. 1956, 12:307-310.

[3] Helen M. Walker and Joseph Lev, Statistical Inference New York: Holt, Rinehart, and Winston, 1953, p. 257.

Counselor variable. Any between counselor differences observed in outcomes of test interpretations can be accounted for by chance. Inasmuch as counselors were selected and trained very carefully in the specific methods of test interpretation utilized, this finding was expected. However, it is encouraging to note that counselors can be trained successfully to do this type of test interpretation.

Relationships of test-estimates and school marks to students' self-estimates. Except for social studies and science at the ninth grade level and science at the eleventh grade level (where correlations between test-estimates and self-estimates were significantly higher than between school marks and self-estimates) chance can account for the observed differences in correlation coefficients. In almost every instance, however, correlation coefficients between school marks and test-estimates were significantly higher than either of the other two correlation coefficients for each subject matter area and grade level.

Relationship of certain variables to father's occupational level. Except for a student's motivation for test interpretation, a father's occupational level does not seem to systematically influence outcomes of test interpretation. Students whose fathers came from higher occupational groups tended to be better motivated for test interpretation. Apparently the greater interest in their children's school achievement exhibited in upper social strata encourages students to request test interpretation.

Comparison of students with highest (top fourth of group) and lowest (bottom fourth) accuracy scores. Students in the top fourth of the group differed significantly from those in the bottom fourth in the following ways: (1) they earned higher mental test scores at all four grade levels; (2) at fifth and ninth grades they earned lower satisfaction scores; (3) they made less increase in accuracy scores after test interpretations; (4) fewer of their fathers come from upper educational levels; and (5) they were motivated to have their tests interpreted but not as highly motivated as their counterparts. An obvious explanation for accurate estimators failure to increase their accuracy scores more than inaccurate estimators is the fact that their pre-test interpretation accuracy level was so high that it left little room for improvement in accuracy scores.

Comparison of students with highest and lowest congruency scores. Students who made the greatest gains in accuracy scores (increased congruency scores) differed from their counterparts as follows: (1) they earned lower mental test scores; (2) more of their fathers had higher levels of education; and (3) they were more highly motivated to have their tests interpreted. These findings support those reported in the two previous paragraphs. It also is interesting that those who have least accurate self-estimates were most highly motivated to have tests interpreted and profited most from having tests interpreted--at least when methods used here were used.

Comparison of accurate estimators with under-estimators and over-estimators. With reference to mental test scores, the under-estimators scored significantly higher in grades nine and eleven than either accurate or over-estimators but the difference between the latter two groups was not significant. For all four grade levels the under-estimators earned the highest mean mental test scores and the over-estimators earned the lowest. The unusually high F ratios obtained in this analysis may be an artifact because the highly intelligent students have greater possibilities for underestimating and the low intelligence students for over-estimating.

With reference to the other variables it was found that: (1) under-estimators were more accurate estimators than over-estimators; (2) over-estimators were most satisfied with their test results--accurate-estimators next most satisfied and under-estimators were least satisfied; (3) over-estimators showed greatest improvement in accuracy scores, and under-estimators profited more from test interpretation than did accurate-estimators; (4) under-estimators had fathers and mothers from higher educational levels and the opposite was true of over-estimators; and (5) all three groups were well motivated to have tests interpreted, but over-estimators were most highly motivated for it. *Under-estimators* were high ability students, were moderately accurate in their initial self-estimates, were well motivated to have tests interpreted, achieved moderate improvements in accuracy scores, and had parents with relatively higher educational attainments. In contrast, *over-estimators* earned the lowest average mental test scores, were least accurate in their initial self estimates, but claimed greatest satisfaction with these estimates, were most highly motivated to have tests interpreted, made the greatest improvement in accuracy scores, and had parents with generally lower level of educational attainments.

Relationship of a teacher's perception of a student's personal appearance, classroom conduct, social and economic background, and academic achievement and discrepancy scores (a teacher's tendency to over-estimate or under-estimate a student's mental test scores). It was assumed that teachers react to students in terms of a general perceptual set and that over-estimation or under-estimation is determined by this perceptual set. If, for example, a child makes a good personal appearance and comes from a good home, the teacher would be inclined to over-estimate his mental test scores. Though a number of the correlation coefficients were large enough to differ significantly from zero, and hence tend to support the notion just presented, only the correlation coefficients between perceptions of academic achievement and discrepancy scores were high enough and consistent enough to offer strong support for the hypothesis tested with reference to this notion. One explanation of this finding may be founded in the idea that a teacher, in attempting to estimate a given student's intellectual ability, may rely heavily upon the student's academic achievement within his classroom. Therefore, as students' classroom achievement goes up, even though it may be a function of such factors as motivation or a highly stimulating home environment rather than basic intelligence, the teacher would see this as indicative of an increasingly better mental ability.

Comparison of discrepancy scores for students rated high and low by their teachers on personal appearance, classroom conduct, social and economic background, and academic achievement. Significant differences in discrepancy scores were obtained for personal appearance, social and economic background, and academic achievement. For those rated low in personal appearance, social and economic background and academic achievement their teachers underestimated their mental test scores and for those rated high in academic achievement their teachers over estimated their mental test scores. This further analysis confirms the results presented in the previous paragraph. If teachers are to have appropriate expectations for their students, they must be aware of this tendency and learn to control it. Perhaps this is less of a problem for the student who is over-rated since he may be motivated to perform at the teacher's expected level, and if he is not, the teacher probably will change his perception to make it conform to his academic achievement level rather than to continue to put pressure on him. For the student whose mental ability is underestimated on the basis of these other characteristics the problem may be considerably more serious. His teacher's evaluation of his mental ability is conveyed to the student causing him to underestimate himself, to set goals which are not educationally challenging for him, and to more readily accept attitudes toward school which are commonly reinforced by his environment. Perhaps this is what often happens to the capable early school leaver.

Extent to which teachers estimate their students mental ability accurately. For most students, teachers do estimate mental ability accurately (correlation coefficient for total groups was .682.). The results obtained in this study support those obtained in other similar studies.

Comparison of teachers' and students' self-estimates of mental ability. Correlation coefficients obtained from teacher estimates and mental test scores were significantly higher than those obtained with students' estimates (.682 vs. .587). However, both were significant and high enough so that they should be obtained and used in making curricular decisions concerning students and in helping students develop educational and vocational plans.

Comparison of students who were overestimated, accurately estimated, and underestimated by their teachers and students own self estimates. First it was discovered that there was a positive relationship between teachers' estimates of students' mental ability and students' self-estimates (.538). Hence, it was not surprising to find that most students who overestimated their mental ability tended to have it overrated by their teachers, and similar results were found for students who underestimated their mental ability and for those who appraised it most accurately.

Finally, these results clearly suggest the students do want to have their tests interpreted and that this method of test interpretation does improve accuracy of self-estimates. Obviously, there is more to interpreting a test than accurately communicating how the test results describe the student. The counselor who interprets a test must know the student and the test which he attempts to interpret. He also must be sensitive to the student's feelings toward what he learns about himself. Apparently some students assimilate and integrate test results quite readily whereas others do not alter their self-images to any appreciable degree.

Perhaps the findings presented here can be used by counselors to better understand students' behavior and to help counselors deal with students reaction to what they learn about themselves. In particular, these results suggest that counselors spend more time with the under-estimators. This group may include many of the early school leavers. Until these students, and their teachers, better understand what may be expected of them, school will not be very meaningful to them and they will not be assigned to tasks in which they can achieve reasonable success.

Further research should be encouraged in all of these problem areas but especially with reference to the way in which a students' social class effects outcomes of test interpretation. To achieve this better data must be obtained for classifying the students' family with reference to social class. These investigators also believe that a student's personal appearance and social-economic background may effect counselors' as well as teachers' reactions to students and their ability to adequately interpret tests. Hence, this problem also should be investigated.

Bibliography

Forty different references were cited in the three papers included in this project report.

VARIABLES RELATED TO OUTCOMES OF TEST INTERPRETATION

Paul A. Wurm and Merle M. Ohlsen
University of Illinois

Purpose of the Study

The increased emphasis upon testing in the public schools in recent years has been accompanied by an increasing interest in evaluating the use of test results. Even today some reasonably well qualified interpreters have not realized that effective test interpretation calls for much more than a knowledge of the test and the ability to communicate what the test results say about students. Perhaps this explains why so few researchers who have evaluated their efforts to increase pupils' self-understanding by interpreting test scores have obtained significant results. Lister and Ohlsen (1962) reported one of the few studies in which pre-college age youth were used as subjects, and increased self-understanding was achieved. They concluded that their results were probably accounted for by two conditions: (1) counselors' efforts to involve clients in test interpretation and (2) counselors' skill in detecting clients' feelings toward test results and in helping them discuss their feelings.

Increased pupil self-understanding should be one of the primary reasons for testing. According to Saper (1961), although counselors interpret tests in their interactions with students, "this use of tests must be subordinated to the wider and more essential aims of counseling. These are to acquaint the student with his assets and liabilities, to help him assimilate this knowledge and act on it, and to be able to make the best possible adjustment."

It is important for every youngster to become better acquainted with his abilities and aptitudes in order for him to determine what he has a right to expect from himself as a student. A pupil also must become increasingly better acquainted with his interests, abilities, and aptitudes in order to define educational and vocational goals and to revise these goals in light of increased knowledge of self and changes in socioeconomic conditions. Therefore both teachers and counselors need information as to what factors seem to be related to the outcomes of test interpretation.

Taking cognizance of phenomenological theory, variables that help to shape a pupil's phenomenological world such as real and ideal self-concepts, perceived performance based on past experiences, motivation to learn more about himself, and socioeconomic status, should have a direct bearing on the extent to which test interpretation increases self-understanding. Based on these grounds, the intention of this study is:

1. To investigate the extent to which outcomes of test interpretations vary with the counselor who interpreted the test scores.

2. To explore the relationships among school marks, self-estimates, and test scores.

3. To examine the relationship between father's occupational group and: (a) accuracy of self-estimates, (b) perceived accuracy of self-estimates, ($_c$) discrepancies between actual and ideal self-estimates, (d) outcomes of test interpretation, and (e) motivation for test interpretation.

4. To compare those students with most accurate self-estimates and those with least accurate self-estimates with reference to: (a) intelligence, (b) perceived accuracy of self-estimates, (c) satisfaction with self-estimates, (d) discrepancies between actual and ideal self-estimates, (e) outcomes of test interpretation, (f) father's occupational group, (g) father's educational level, (h) mother's educational level, and (i) motivation for test interpretation.

5. To compare those students with the greatest gains in accuracy over test interpretation and those with the least gains in accuracy with reference to: (a) intelligence, (b) perceived accuracy of self-estimate, (c) satisfaction with self-estimates, (d) discrepancies between actual and ideal self-estimates, (e) father's occupational group, (f) father's educational level, (g) mother's educational level, and (h) motivation for test interpretation.

6. To compare those students with most accurate self-estimates against both over-estimators and under-estimators with reference to: (a) intelligence, (b) perceived accuracy of self-estimates, (c) satisfaction with self-estimates, (d) discrepancies between actual and ideal self-estimates, (e) outcomes of test interpretation, (f) father's occupational group, (g) father's educational level, (h) mother's educational level, and (i) motivation for test interpretation.

Review of Related Research

Many conscientious workers present test scores, but often fail to improve students' understanding of themselves. Though some studies clearly indicate that careful interpretation of test scores by qualified personnel does improve students' self-knowledge, there are other studies which indicate failure to produce increased self-understanding. Those studies which used high school students as subjects will be reviewed first, followed by those which used college students as subjects. Except for the Lister and Ohlsen (1962) study, there is nothing to report on the extent to which interpreting test scores improves accuracy of self-image for elementary school children.

Adamek (1961) compared changes in perceptions of interests, mental abilities, and problem areas of ninth graders. He used two experimental groups and one control group in a testing program which involved Kuder Preference Record - Vocational, Chicago Primary Mental Abilities, and the SRA Youth Inventory. For one group, the tests were interpreted by a well qualified counselor in an individual interview in which emphasis was placed upon communication of information, and no effort was made to deal with affect. For the other group the same counselor used the publisher's suggested technique for self interpretation in small groups. He concluded that:

> there were no systematic differential effects by which one type of test interpretation could be concluded to be superior to the other.

Froehlich and Moser (1954) also used ninth-graders as subjects. They administered the Differential Aptitude Test Battery, had pupils draw their own profiles during test interpretation, and permitted them to keep their profiles. In a fifteen month follow-up the pupils were asked to draw their profiles from memory. Most students did not report their scores accurately in the follow-up. Both high and low ranking pupils made errors in reporting; however, a larger proportion of brighter pupils reported scores more accurately than did the less bright.

Lallas (1956) studied high school juniors, comparing three methods of test interpretation: individual interview, group interpretation, and group interpretation plus individual interview. The results showed that after test interpretation all three experimental groups were superior to the control group in accuracy of estimating their ability, and in identifying areas of greatest and least ability.

Belovsky, McMaster, Shorr, and Singer (1953) used twelfth-graders as subjects. The treatment consisted primarily of testing and individual interpretations of the tests for one group, and of testing and group interpretation of the tests for the other. Using realism of vocational choice as a criterion, they found that chance could account for the slight differences noted between the two groups.

Lane (1952) studied eleventh and twelfth-graders to compare the effectiveness of "traditional" and "permissive" techniques of interpreting test results. He found almost identical scores for the two groups and hence concluded that there were no differences in the effectiveness of the methods.

In a study using high-school students, Singer and Stefflre (1954) obtained self-estimates of interests prior to testing and counseling, and three months following test interpretations. They used discrepancy between self-estimate and test-estimate as a measure of accuracy. They found no significant differences between means; however, the variance of the discrepancy scores post-counseling was significantly smaller which suggested increased congruency between test-estimates and self-estimates following counseling.

Biersdorf (1958) recruited male college students who were interested in receiving assistance with vocational plans. Using random methods the students were divided into three groups: (1) limited treatment group,

(2) extended treatment group, and (3) control group. The limited treatment group received group interpretation of the vocationally relevant tests which they had taken. The extended treatment group received, in addition to the test interpretation, a discussion of factors important in making vocational plans. She concluded that neither was more effective than no treatment.

Dressel and Matteson (1950) noted that there were wide variations in procedures used by their counselors in interpreting tests. These procedures differed mainly in the amount of client participation involved. Using university freshmen as clients they found that the mean gains in self-understanding made by clients appeared to be closely related to the mean client participation index. They concluded that counselors who encouraged a high degree of client participation were more successful in stimulating client growth than were the other counselors.

Gustad and Tuma (1957) investigated three methods of test introduction in which the counselor assumed varying degrees of responsibility for introducing the idea of testing. They also used four methods of test interpretation in which counselors assumed varying degrees of responsibility for pointing out discrepancies between students' self-estimates and test scores. These methods of test introduction and test interpretation were compared as to their effect on client learning. No significant differential effects were noted among the various methods. They did find, however, that client learning was positively related to initial accuracy of self-ratings, but not to scholastic aptitude. In another paper (Tuma and Gustad, 1957) they investigated the effects of client and counselor personality on client learning. Counselors who used essentially the same method with similar clients produced different effects. The investigators concluded that there was a positive relationship between amount of client learning and amount of client-counselor similarity.

Holmes (1961) used college freshmen to compare the effectiveness of four techniques of presenting test information. The four techniques used were as follows: (1) The counselor described each test and pointed out the student's standing on this profile, and then answered questions. (2) The counselor asked each student to estimate his performance on a test profile before presenting actual results, and allowed the client to choose the order of test interpretation. The counselor maintained a non-evaluative attitude, but concentrated on the effect the student expressed regarding the test results and self-estimates. (3) The counselor selected the order of test interpretation and presented the results. Student participation was encouraged by exploring, clarifying, and reflecting students' feelings and attitudes toward the test results. (4) Students were mailed a brief summary of the meaning of each test and a test profile on which the scores were plotted.

Holmes concluded that in all four methods of test interpretation the test interpretation process appeared useful to students in helping them to understand themselves, their vocational choices, interests, and strengths and weaknesses; but that students who received their results by mail considered them less valuable. The method used by a counselor had less influence on student attitudes toward the counselor than the personality of the counselor himself.

Kamm and Wrenn (1950) appraised client's acceptance of self-information by studying educational-vocational planning interviews with college freshmen. They reported that a client was most apt to accept information when both the client and the counselor were relaxed, when the client was expressing positive attitudes, when the information which was given was directly related to the client's immediate problem, and when the information was not in opposition to the client's self-concept. They concluded that a counselor would do well to pay more attention to how the client seems to feel about what he is learning about himself, and devote less attention to the content of what the client says.

L. B. Rogers (1954) compared two kinds of test interpretation with college students: (1) a test-centered method in which there was little student involvement, and (2) a method of self-evaluation where great emphasis was placed on self-study and use of non-test data in addition to test results. Both methods seemed to increase students' self-understanding with respect to abilities and interests; however, there were no statistically significant differences in effectiveness between methods.

Torrence (1954) used college freshmen as subjects in a study of accuracy of self-estimate. He found little relationship between self-estimates and test scores, however, when students were asked to re-evaluate themselves after having conferences with counselors and academic advisors, they generally revised their estimates in the direction of more realistic self-evaluation.

Wright (1957) in a study on entering freshmen compared the effectiveness of test interpretation conducted in individual interviews with test interpretation conducted in group situations. The members of the two experimental groups were invited by letter to make an appointment to review the results of their freshmen guidance examinations. Members of one group received their test results in individual test interpretation interviews. Members of the other group met in small groups of five to ten students to review their test results. The groups were comparable in terms of age, sex, intelligence, and high school grade-point average. Following counseling, both groups showed significant increases in accuracy of self-ratings of aptitude and achievement.

Johnson (1953) studied voluntary male guidance center clients using a self-rating device and found that vocational counseling increased accuracy and certainty of self-knowledge. Gains remained stable after a one-month follow-up. Gains in self-understanding were greatest with respect to intelligence, followed by those with respect to interest and personality, respectively.

Lister and Ohlsen (1962) attempted to select counselors with care and to prepare them to involve students in test interpretation and to take account of how students seem to feel about what they learn about themselves. By using fifth, seventh, ninth, and eleventh-grade students they also attempted to assess the extent to which maturity is a factor which effects outcomes of test interpretation. They investigated the effects of pretesting orientation upon students' motivation for learning test results, and the changes in accuracy of self-estimates resulting from test interpretation. They found that (1) students, at all four grade levels, increased the accuracy of their self-estimates following test interpretation and maintained a significantly more accurate self-estimate over a two-month follow-up period; (2) the orientation resulted in greater motivation for learning test results within grades seven and nine, but not within grade five and eleven (approximately nine out of ten of all upper grade students wanted their tests interpreted in both the orientation and control groups); (3) there was no evidence of a relationship between orientation and improved accuracy of self-estimates following test interpretation; (4) there was support for the hypothesis of a relationship between motivation for test interpretation and increased accuracy of self-estimates following test interpretation; (5) there was a significant positive correlation (though relatively low) between measured intelligence and students' accuracy of self-estimates of achievement and intelligence, but a negative one between measured intelligence and students' ability to correctly estimate their interests; and (6) students' satisfaction with self-estimates and perceived accuracy of self-estimates increased over the four self-rating periods. The authors suggest that this indicates increased acceptance of self-information.

In summary, it is possible to infer from these studies that changes in self-perceptions do occur as a result of test interpretation, although there is no clear-cut evidence as to which of the many methods used is the more effective in producing these changes. It is also possible to infer that maturity may be a factor in determining the extent to which students profit from test interpretation. It would also seem that counselors would do well to attempt to involve students actively in the test interpretation process, and to pay special attention to students' feelings about the things they learn about themselves and help them deal with these feelings.

The majority of the studies reviewed dealt mainly with differing methods of test interpretation procedures and the effects of these precedures along with other test factors and counselor variables on the outcomes of test interpretation. Primarily these studies were done to determine the relative effectivenss of the different methods used. No study reviewed set about examining, in any systematic manner, the socioeconomic or experiential factors in students' backgrounds which may effect self-estimations and outcomes of test interpretation.

The present study attempts to extend the investigation of the test interpretation process and its effects upon students by examining some of the above mentioned factors that may have relevance to increased self understanding in students.

Definition of Terms

Significant, as used in the present study, refers to statistical results that have a probability of equal to or less than .05 of being due to the operation of chance factors alone if there were indeed no relationship present in the population.

A student's Accuracy Score refers to the differences between his self-estimated test scores and his actual test scores.

A student's Congruency Score refers to his change in accuracy from pre to post-test interpretation.

Procedures

The Sample

The majority of the data used in the present investigation were collected under Cooperative Research Project #1344, U. S. Office of Education, conducted by Lister and Ohlsen (1962). Included were scores from a standardized test battery, and scores on a Self-Rating Scale. Since that time the senior author collected the school marks of the pupils involved, and classified students' father's occupational group, father's educational level, and mother's educational level.

Five very similar school districts within a twenty mile radius of the University of Illinois participated in the original study. They were rural communities in which also lived a good cross-section of all occupational levels, with perhaps slightly more professional people than the typical Illinois towns of similar size. Although it was necessary to test all 1,578 fifth, seventh, ninth, and eleventh-graders in order to secure cooperation of these school districts, only 760 students were used in the original study. Of those 760 who participated, it was possible to collect the complete data necessary for the present investigation on a total of 703. Distribution by grade was: fifth-grade - 190; seventh-grade - 176; ninth-grade - 204; eleventh-grade - 133.

Instruments

The instruments used consisted of a battery of standardized tests, the Self-Rating Scale, the Counseling Request Form, and the Information Sheet.

The standardized tests selected for administration in the original study were representative of test batteries commonly given in public schools. They included measures of achievement, intelligence, and interests.

In grades five and seven, the Iowa Tests of Basic Skills, Form 1, were administered in late October and early November. These tests yield scores on vocabulary, reading, language, work-study skills, and arithmetic.

While attention must be given to specific content in any adequate achievement battery, the Iowa Tests of Basic Skills have emphasized generalized understandings rather than recall of specific bits of information.

The Iowa Tests of Basic Skills are among the most reliable achievement tests available. The reported split-half reliabilities range from .84 to .96 for the tests and from .97 to .98 for the composite score. Since these data were obtained from tests administered at the beginning of the school year, one would expect mid-year and end-of-year reliability coefficients to be even higher (Herrick, 1959).

In grades nine and eleven, the Iowa Tests of Educational Development, Form X3-S, were administered. Of the nine tests available in the class period length, the following were administered: Test 1, Background in the Social Sciences; Test 2, General Background in the Natural Sciences; Test 3, Correctness and Appropriateness of Expression; Test 4, Ability to do Quantitative Thinking; and Test 8, General Vocabulary.

As in the case of the Iowa Tests of Basic Skills, one of the reasons given by Lister and Ohlsen (1962) for selecting these tests was "that the items were designed to sample understanding and ability to apply concepts rather than the recall of specific information, per se".

"The odd-even reliability coefficients for these tests ranged from .84 on Test 4 to .94 on Test 8 with a coefficient of .98 for the composite score for grade nine. In grade eleven, the odd-even reliabilities for the tests ranged from .87 on Tests 2 and 3 to .95 on Test 8 with a coefficient of .98 for the composite score. (Obtained by writing to the publisher.)" (Lister and Ohlsen, 1962).

The Henmon-Nelson Tests of Mental Ability, Form A, were administered to all grades. "These tests were selected because they provided reasonably reliable and valid measures of general intelligence with a minimum of testing time." (Lister and Ohlsen, 1962)

The median test-retest reliability coefficient reported is .94 with values ranging from .90 to .97 (Lefever, 1959).

The reported validity evidence for the Henmon-Nelson Tests was considered highly satisfactory by Lister and Ohlsen (1962). The median coefficient for total achievement battery scores and I.Q.'s is .79 with a range from .64 to .85. Average grades and I.Q. yielded a median correlation coefficient of .60 with a range from .09 to .74. Correlations between I.Q.'s and achievement subtests for various grade levels ranged from .65 to .76. Median correlations between I.Q.'s and teachers' grades for various grade levels ranged from .53 to .71 (Lefever, 1959).

In grades five and seven, the What I Like to Do: An Inventory of Children's Interests (Bonsall, et. al., 1954) was administered. The Kuder Vocational Preference Record, Form C (1960), was administered in grades nine and eleven.

Due to the known instability of students' interest scores, especially at the lower grade levels, and the apparent inability of students to accurately estimate their interest scores in the original study, (Lister and Ohlsen, 1962), interest test data were not utilized in the present investigation.

The Self-Rating Scale used in the original study was designed in two forms: one form for grades five and seven and another form for grades nine and eleven. The rating scales attempted to describe, in language appropriate to the grade level, the areas measured by the respective tests. The rating scales were designed to provide the following kinds of information:

(1) Students' present estimates of their performance on the respective subtests of the test battery. Students were instructed to indicate this estimate by checking one of five statements from "Very much more than most students my age" to "Very much less than most students my age."

(2) Students' current satisfaction with their estimated performance. Students were instructed to indicate this satisfaction by checking one of five statements from "I very much like being this way" to "I very much dislike being this way."

(3) Students' current ideal performance on the respective subtests of the test battery. Students were instructed to indicate this ideal performance by checking one of five statements from "Very much more than most students my age" to "Very much less than most students my age."

(4) Students' estimated accuracy of their self-ratings of the respective test areas of the test battery. Students were instructed to indicate the estimated accuracy of their self report by checking one of five statements from "I know almost exactly how I compare with others" to "I really don't have any idea about how I compare with others."

Considering the reliability of the Self-Rating scale, Lister and Ohlsen (1962) reported that the "median (test-retest) correlations for estimates of actual performance, satisfaction with actual performance, and ideal performance for all grades were approximately .60."

The Counseling Request Form was designed to secure students' responses to the opportunity to participate in individual test-interpretation, counseling interviews. On completion of the test battery students were asked to check one of three statements on the request form:

1. I would like to talk about my test results with a counselor to learn more about myself.

2. I am not sure at this time whether I want to talk about my test results with a counselor to learn more about myself. I will decide about this in a few days.

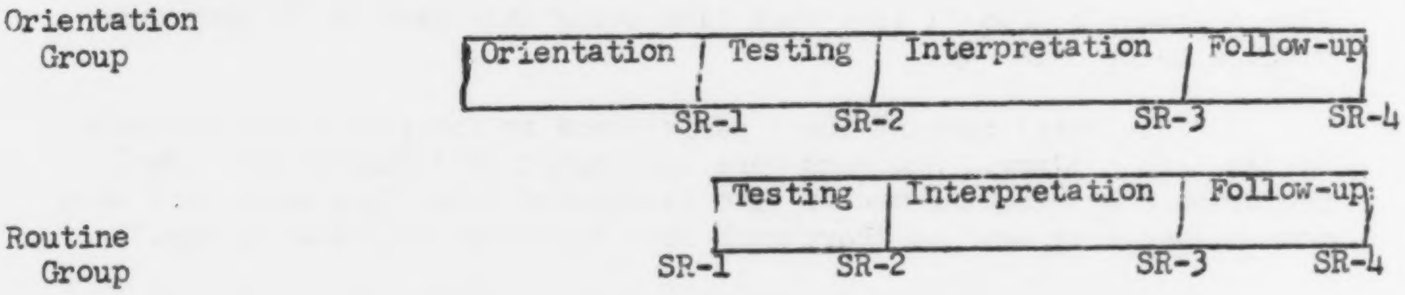
3. I would rather not talk about my test results with a counselor to learn more about myself.

Students who checked the second statement were given one week to make a decision regarding participation in the test interpretation interview. At the end of the week, forms were returned to these students, and they were asked to check either the first or third statement.

Data on the father's occupation, father's education, and mother's education were obtained from an information sheet which was sent home with children to be completed by parents. This information was collected at the time of the original study, but was not analyzed at that time.

Method of Data Collection

The experimental design of the original study was of the following nature:



As Lister and Ohlsen (1962) found no significant differences between the Orientation and Routine groups, they were pooled for the purpose of the present investigation. For the current study, SR-2 was used as the pre-counseling self-rating, and SR-4 was used as the post-counseling self-rating. It was felt that using SR-4 rather than SR-3 as the post-measure would eliminate some of the rote memory factor which might be present as a result of the self-rating scales being filled out immediately following counseling.

The Achievement and Intelligence test results used in the present investigation came from the testing phase of the original study. Father's occupation, and father's and mother's education data came from the information Sheet which was administered at the end of the testing phase. Also administered at that time was the Counseling Request Form, from which the student's motivation was assessed.

The data obtained from the Information Sheet administered in the original study were classified by the present investigator during the Fall term of the 1963-64 school year. The father's occupational group was determined by classifying eachefather's occupation according to a modified Dictionary of Occupational Titles classification including the following categories: (1) professional and managerial occupations; (2) clerical and sales occupations, (3) service occupations: (4) farm operators and managers; (5) skilled occupations; and (6) semi-skilled and unskilled occupations, including farm laborers.

The father's and mother's educational levels were classified by the following system: (1) 0-6 years of formal education; (2) 7-8 years of formal education; (3) 9-10 years of formal education; (4) 11-12 years of formal education; (5) 13-14 years of formal education; (6) 15-16 years of formal education; and (7) beyond 16 years of formal education.

During the Fall term of the 1963-64 school year, the present investigator returned to the cooperating schools and collected certain school marks earned by students during the first semester of the year in which the original project was carried out. The Fall term marks were collected because that was the last grading period occurring prior to test interpretation. The marks collected for grades five and seven were: Reading, Language, and Arithmetic. For grades nine and eleven: Social Studies, Science, English, and Mathematics marks were collected.

Data

The raw data obtained for use in the present study were of two basic types. The first type consisted of qualitative variables including father's occupational group (six categories), father's educational level (seven categories), mother's educational level (seven categories), and motivation (two categories: requested test interpretation or did not request test interpretation). The second type consisted of continuous data including test scores, self-ratings, and school marks.

From the continuous variable data, several derived scores were computed for use in testing the specific hypotheses. These derived scores were: (1) Accuracy scores; (2) Perceived accuracy scores; (3) Discrepancy scores; (4) Satisfaction scores; and (5) Congruency scores. These scores were arrived at in the following manner.

1. Accuracy scores.

In the test interpretation a student's test score was indicated on the interview form. For each subtest of the battery, the score was indicated on a continuum divided into five bands approximating the five choices given on the Self-Rating scale. To determine a student's overall accuracy score, the difference between the self-estimate and the test-score was subtracted from a constant (five) for each subtest, and these differences summed across the five achievement subtests and the intelligence test. This process yielded scores on a continuum the high end of which indicates high accuracy.

2. Perceived Accuracy scores.

On the Self-Rating scle, students were asked to indicate on a five point scale how accurately they felt they had estimated themselves in each of the three areas of Achievement, Intelligence, and Interests. For the present investigation, a student's perceived Accuracy score was computed by summing his self-ratings for the areas of Achievement and Intelligence. This yielded scores on a continuum the high end of which indicates low perceived accuracy.

3. Discrepancy scores.

A part of the Self-Rating scale asked the students to rate themselves on each subtest, as to how they would really like to be. Discrepancy scores were computed by summing the absolute differences between this "ideal" self-rating and the "real" self-rating across all six subtests. This process yielded scores on a continuum the high end of which indicates high discrepancy.

4. Satisfaction scores.

Another task presented to students on the Self-Rating scale was to indicate, on a five point scale, how they felt about being the way they described themselves in their "real" self-estimates. Satisfaction scores were arrived at by summing this self-rating across all six subtests. This yielded scores on a continuum the high end of which indicates low satisfaction.

5. Congruency scores.

A student's Congruency score refers to his change in accuracy from pre-to post- test interpretation. Congruency scores were determined by subtracting a student's Accuracy score pre-interpretation from his Accuracy score post-interpretation, and adding this difference to a constant (twenty-five). This process yielded scores on a continuum where a score of 25 indicates no change in accuracy, scores higher than 25 indicate gains in accuracy, and scores below 25 indicate losses in accuracy. In general, higher scores indicate more desirable outcomes of test interpretation.

The qualitative data, test scores, self-ratings, school marks, derived scores, and identifying information were all punched into I.B.M. cards, and all data analyses were carried out utilizing the I.B.M. 1401 computer at the University of Illinois Statistical Service Unit.

Results

In the present investigation, ten major hypotheses were tested. Comparisons were made among various specific sub-samples of the participating students on a number of variables. In general, statistical tests made on these sub-samples were carried out separately for each of the four grade levels and for the total sample. In the presentation of results each hypothesis is stated, followed by the statistical procedures used to test the hypothesis including the organization of raw data, and the results and discussion of the hypothesis. In those cases where the hypothesis consists of several specific sub-hypotheses, a discussion of the hypothesis as a whole follows the presentation of the results of all of the sub-hypotheses.

Hypothesis 1.

When students' Congruency scores are compared on the basis of the counselor who interpreted their test scores, no significant difference exists among mean congruency scores of the groups.

Statistical Procedures for Hypothesis 1.

To test hypothesis 1, students were classified into treatment levels based on the counselor who interpreted their test scores. A single classification Analysis of Variance was then performed on the Congruency scores earned by these students.

Results and Discussion of Hypothesis 1.

The results shown in Table 1 indicate that the decision to retain this hypothesis is warranted. Apparently none of the twenty counselors who interpreted tests to students did a singularly better job than the others, at least in terms of students' changes in accuracy as reflected by Congruency scores.

TABLE 1

ANALYSIS OF VARIANCE OF CONGRUENCY SCORES OF STUDENTS SEEN BY DIFFERENT COUNSELORS

Sample	Source of Variation	Sum of Squares	df	Mean Square	F	P
Grade 5	Between	71.414	11	6.492	1.198	NS
	Within	963.960	178	5.415		
	Total	1035.374	189			
Grade 7	Between	82.058	14	5.861	1.015	NS
	Within	929.579	161	5.773		
	Total	1011.637	175			
Grade 9	Between	68.618	13	5.278	1.111	NS
	Within	901.971	190	4.747		
	Total	970.589	203			
Grade 11	Between	47.566	15	3.171	0.871	NS
	Within	425.592	117	3.637		
	Total	473.158	132			
Total Sample	Between	113.255	19	5.960	1.198	NS
	Within	3396.541	683	4.972		
	Total	3509.796	702			

This hypothesis was tested as a type of control. Since several studies (Dressel and Mattleson, 1950; Tuma and Gustad, 1957; and Holmes, 1961.) reported effects from the counselor variable, it was deemed necessary that this hypothesis be supported before completing other analyses involving Congruency scores. If no support had been found for this hypothesis, it would have been necessary to assume that some counselors performed better than others, and would have necessitated controlling for this fact in all further analyses of Congruency scores.

It is hardly surprising that support was found for this hypothesis in light of the fact that the counselors selected for interpreting tests in the original project were carefully chosen, and well trained in the specific method of test interpretation utilized. For further information on this method see Chapter 5 in *The Effects of Orientation to Testing on Motivation for and Outcomes of Test Interpretation.* (Lister and Ohlsen, 1962)

Hypothesis 2.

Prior to test interpretation the relationship between self-estimate and test scores will not be significantly higher than the relationship between self-estimate and school marks.

Statistical Procedures for Hypothesis 2.

To test this hypothesis, product-moment correlations were computed between self-estimates made prior to test interpretation and school marks, and between those same self-estimates and test scores in each of the achievement areas. The obtained correlations were then compared utilizing the procedures outlined in Walker and Lev (1953, p. 257). This procedure yields a *t* ratio with N-3 degrees of freedom.

Results and Discussion of Hypothesis 2.

As can be seen from Table 2, the results tend to support this hypothesis. No significant differences were obtained between the correlations with the exception of three cases, and in these three cases the direction of significance was opposite from that which was expected.

The underlying rationale of hypothesis 2 was the assumption that students would be more apt to rate their achievement on the basis of past experience as reflected by school marks than on the basis of how well they thought they did on specific standardized achievement tests. The results of the present study would seem to indicate that this assumption is unwarranted.

One apparent alternative to this assumption is that students are able to adequately distinguish between class grades and performance on standardized achievement tests. This would seem to be unlikely at the lower grade levels. It is interesting to note that, in general, the correlations tend to become higher as grade level increases. Perhaps due to increased cognitive sophistication, the older age-groups have greater ability to perform the required task of estimating test scores.

TABLE 2

PRODUCT-MOMENT CORRELATIONS AMONG SELF-ESTIMATES TEST SCORES, AND SCHOOL MARKS

		Correlations				
		A	B	C		
Sample	Achievement Area	Self-Estimate vs. Test Mark	Self-Estimate Vs. School Mark	School Mark Vs. Test Score	± (A-B̄)	P
Grade 5 (N 190)	Reading	.224**	.242**	.615**	-0.291	NS
	Language	.168*	.231**	.566**	-0.951	NS
	Arithmetic	.156	.234**	.610**	-1.245	NS
Grade 7 (N 176)	Reading	.227**	.251**	.701**	-0.423	NS
	Language	.261**	.254**	.724**	0.129	NS
	Arithmetic	.271**	.226**	.658**	0.746	NS
Grade 9 (N 204)	Social Studies	.393**	.225**	.451**	2.481	<.01
	Science	.375**	.267**	.483**	1.701	<.05
	English	.164*	.217**	.589**	-0.852	NS
	Mathematics	.348**	.294**	.314**	0.715	NS
Grade 11 (N 133)	Social Studies	.463**	.423**	.656**	0.630	NS
	Science	.531**	.297**	.570**	3.395	<.01
	English	.277**	.392**	.562**	-1.531	NS
	Mathematics	.435**	.359**	.545**	1.025	NS

**Significant at or beyond α = .01
*Significant at or beyond α = .05

It should be pointed out that in the present investigation school marks were distributed on a five point scale as were self-estimates and test scores. Correlations computed on the basis of these relatively broad classification categories would not take advantage of the variation present within these categories, especially in the case of self-estimates and test scores.

Finally, it can be noted from Table 2 that the relationships between school marks and test scores were relatively high. Correlations ranged from .31 to .72 with a median of .58. In other words, if both the test score and the school mark variables are reflections of the student's ability, and the student estimates his ability with a fair amount of accuracy, it is not surprising that no significant differences exist between the correlations of self-estimates with school marks, and the correlations of self-estimates with test scores.

Hypothesis 3.

When students are classified on the basis of father's occupational group, no significant differences exist among mean:

a. Accuracy scores of the groups.
b. Perceived Accuracy scores of the groups.
c. Satisfaction scores of the groups.
d. Discrepancy scores of the groups.
e. Congruency scores of the groups.

Statistical Procedures for Hypothesis 3.

For hypothesis 3, students were classified on the basis of their father's occupational group which was determined from the father's occupation as listed on the Information Sheet collected at the time of testing. Using these as the treatment groups, a single classification analysis of variance was computed for each of the variables specified in the sub-hypotheses. In those cases where significant F ratios were obtained, Kramer's extension of Duncan's new multiple range test (Kramer, 1956) was applied to make multiple comparisons among the group means.

Results and Discussion of Hypothesis 3.

3a. Differences among Accuracy scores of the six occupational groups.

The results shown in Table 3 indicate that the decision to retain this hypothesis at all grade levels and Total N is warranted. Apparently students from the six different father's occupational groups are homogeneous with respect to accuracy of self-estimation, and this homogeneity is consistent throughout the grade levels.

3b. Differences among Perceived Accuracy scores of the six occupational Groups.

The results shown in Table 4 indicate that this hypothesis should be retained within any specific grade level, but rejected when the analysis is

TABLE 3

ANALYSIS OF VARIANCE OF ACCURACY SCORES OF STUDENTS FROM SIX FATHER'S OCCUPATIONAL GROUPS

Sample	Source of Variation	Sum of Squares	df	Mean Square	F	P
Grade 5	Between	25.573	5	5.114	0.493	NS
	Within	1905.501	184	10.355		
	Total	1931.074	189			
Grade 7	Between	22.949	5	4.589	0.591	NS
	Within	1319.961	170	7.764		
	Total	1342.910	175			
Grade 9	Between	22.774	5	4.554	0.646	NS
	Within	1395.633	198	7.048		
	Total	1418.407	203			
Grade 11	Between	28.367	5	5.673	0.889	NS
	Within	809.724	127	6.375		
	Total	838.091	132			
Total Sample	Between	17.637	5	3.527	0.437	NS
	Within	5620.381	697	8.063		
	Total	5638.018	702			

TABLE 4

ANALYSIS OF VARIANCE OF PERCEIVED ACCURACY SCORES OF STUDENTS FROM SIX FATHER'S OCCUPATIONAL GROUPS

Sample	Source of Variation	Sum of Squares	df	Mean Square	F	P
Grade 5	Between	27.254	5	5.450	2.053	NS
	Within	488.457	184	2.654		
	Total	515.711	189			
Grade 7	Between	21.436	5	4.287	1.662	NS
	Within	438.474	170	2.579		
	Total	459.910	175			
Grade 9	Between	8.727	5	1.745	1.073	NS
	Within	321.876	198	1.625		
	Total	330.603	203			
Grade 11	Between	6.081	5	1.216	0.626	NS
	Within	246.446	127	1.940		
	Total	252.527	132			
Total Sample	Between	32.871	5	6.574	2.994	$<.05$
	Within	1530.310	697	2.195		
	Total	1563.181	702			

performed on the total sample. Evidently students from various father's occupational groups appear to be homogeneous in terms of Perceived Accuracy of self-estimation within a given grade level, with respect to a yes-no statistical decision; but a trend is established which becomes significant when the total sample is analyzed. In the total sample, students in the higher father's occupational groups tend to perceive themselves as more highly accurate in self-estimation. When group means were ordered from lowest to highest (6, 3, 5, 4, 1, 2) the multiple range test revealed significant differences between group 2 and group 5, between group 2 and group 3, and between group 2 and group 6. No other comparisons were significantly different.

TABLE 5

ANALYSIS OF VARIANCE OF SATISFACTION SCORES OF STUDENTS FROM SIX FATHER'S OCCUPATIONAL GROUPS

Sample	Source of Variation	Sum of Squares	df	Mean Square	F	P
Grade 5	Between	168.084	5	33.616	2.362	< .05
	Within	2618.595	184	14.231		
	Total	2786.679	189			
Grade 7	Between	61.055	5	12.211	0.829	NS
	Within	2503.673	170	14.727		
	Total	2564.728	175			
Grade 9	Between	27.870	5	5.574	0.555	NS
	Within	1987.714	198	10.038		
	Total	2015.584	203			
Grade 11	Between	19.409	5	3.881	0.375	NS
	Within	1311.840	127	10.329		
	Total	1331.249	132			
Total Sample	Between	39.527	5	7.905	0.567	NS
	Within	9714.303	697	13.937		
	Total	9753.830	702			

3c. Differences among Satisfaction scores of the six occupational groups.

The results shown in Table 5 seem to warrant the decision to reject this hypothesis within the fifth grade level, but to retain it within the seventh, ninth, and eleventh grade levels, and in the total sample. It would seem that fifth grade students from the higher father's occupational groups tend to be more highly satisfied with their self-estimates. In the fifth grade sample, the multiple range test revealed that the only significant difference was between the two extreme groups, group 3 and group 6. No other comparisons were significantly different. The ordering of the group means from lowest to highest was: 6, 4, 1, 5, 2, 3. In the seventh, ninth, and eleventh grades, and in the total sample, students from the various father's occupational groups seem to be homogeneous with respect to satisfaction.

3d. Differences among Discrepancy scores of the six occupational Groups.

The results shown in Table 6 indicate that this hypothesis should be retained within each of the fifth, seventh, ninth, and eleventh grade levels, and in the total sample. Apparently students from the various father's occupational groups are homogeneous with respect to disparity between real and ideal self-estimates, and this homogeneity is consistent throughout grade levels.

TABLE 6

ANALYSIS OF VARIANCE OF DISCREPANCY SCORES OF STUDENTS FROM SIX FATHER'S OCCUPATIONAL GROUPS

Sample	Source of Variation	Sum of Squares	df	Mean Square	F	P
Grade 5	Between	48.727	5	9.745	0.788	NS
	Within	2274.542	184	12.361		
	Total	2323.369	189			
Grade 7	Between	19.363	5	3.872	0.494	NS
	Within	1330.086	170	7.824		
	Total	1349.449	175			
Grade 9	Between	28.947	5	5.789	0.867	NS
	Within	1321.053	198	6.671		
	Total	1350.000	203			
Grade 11	Between	25.439	5	5.087	0.902	NS
	Within	716.080	127	5.638		
	Total	741.519	132			
Total Sample	Between	12.217	5	2.443	0.292	NS
	Within	5821.314	697	8.351		
	Total	5833.531	702			

3e. Differences among Congruency scores of the six occupational Groups.

On the basis of the results shown in Table 7, the decision to retain this hypothesis appears to be warranted. This hypothesis is retained at all grade levels and in the total sample. Evidently the tendency to gain in accuracy as a result of test interpretation is not a function of students' father's occupational group.

The underlying purpose of hypothesis 3 was to examine the influence of socioeconomic status, as reflected in father's occupational group, upon the variables of accuracy and perceived accuracy of self-estimates, satisfaction with self-estimates, discrepancy between "real" and "ideal" self-estimates, and outcomes of test interpretation in terms of changes in accuracy. It was felt that, due to the known differences in attitude toward educational achievement among the social class strata in American society, the variable of socioeconomic status would have an effect upon students' perceptions toward and feelings about their ability and achievement.

TABLE 7

ANALYSIS OF VARIANCE OF CONGRUENCY SCORES OF STUDENTS FROM SIX FATHER'S OCCUPATIONAL GROUPS

Sample	Source of Variation	Sum of Squares	df	Mean Square	F	P
Grade 5	Between	46.110	5	9.222	1.715	NS
	Within	989.264	184	5.376		
	Total	1035.374	189			
Grade 7	Between	22.880	5	4.576	0.786	NS
	Within	988.757	170	5.816		
	Total	1011.637	175			
Grade 9	Between	14.412	5	2.882	0.596	NS
	Within	956.177	198	4.829		
	Total	970.589	203			
Grade 11	Between	9.519	5	1.903	0.521	NS
	Within	463.639	127	3.650		
	Total	473.158	132			
Total Sample	Between	22.428	5	4.485	0.896	NS
	Within	3487.368	697	5.003		
	Total	3509.796	702			

The results obtained in the present investigation do not yield clear-cut evidence for the support or rejection of this point of view. If this factor were operating, one would expect that parental and societal pressure would cause students in the upper father's occupational groups (middle-to upper middle class) to hold attitudes that would lead them to be more selectively attentive to their ability and achievement, and would result in their being more accurate in their self-estimates and perceive themselves as being more accurate. In the present study, no evidence was found to suggest that students from higher occupational groups were indeed more highly accurate. In the case of perceived accuracy, however, there was evidence to suggest that those students with fathers in higher occupational groups tended to perceive themselves as more accurate.

Perhaps the effect of the factor in question is opposite, in a way, from what one would expect. In other words, it may be that the very pressures exerted on children from middle class environments cause them to be concerned enough about their ability and achievement that they tend to estimate themselves unrealistically. The results of hypothesis 10a of the present study do not provide evidence of a statistically significant direction (i.e. under or over-estimation) of error. Apparently the situation is one of individual variation. One student may under-estimate himself out of a basic reluctance to admit his own ability for fear of not being able to live up to it in the future, or perhaps because he does not wish to admit to his peers or others that he feels he is superior; whereas another might over-estimate his ability out of a desire to be more able and accomplished than he really is. This situation is one which should be carefully examined in any future study similar in nature to the present investigation.

The situation is also unclear in terms of satisfaction with self-estimates. Students from the fifth grade level in the higher occupational groups appear to be more highly satisfied. At the other grade levels, however, this trend is not present. Satisfaction would appear to be a complex variable. In all probability it is dependent not so much on social class factors as on a combination of such things as "ideal" self and accuracy of estimation, in addition to attitudes toward educational achievement. There seems to be no evident reason for the relationship being present only at the fifth grade level, unless students from higher occupational groups are generally more highly satisfied with school and school experiences at this level, and this general satisfaction carries over to self-estimation. This could be true due to the fact that these students are generally more verbal, bright, and neat appearing than their counterparts, and thus get more attention and approval from teachers, in addition to receiving more attention and approval in regard to school work from parents. This would not necessarily continue to be true throughout grade levels, for as children mature they tend to shift their locus of evaluation more and more away from home and teachers, and towards the peer culture.

Evidently discrepancy between "real" and "ideal" self-estimates, and outcomes of test interpretation in terms of changes in accuracy are not related to social class as evidenced by father's occupational group. No significant differences among occupational groups were found, and no consistent trends were noted in these variables.

In terms of the outcome variable, this would seem to indicate that, in general, test-interpretation counseling is equally valuable and effective with students from all class levels.

Hypothesis 4.

No significant relationship exists between father's occupational group and student's motivation for test interpretation.

Statistical Procedures for Hypothesis 4.

The father's occupational group classification used in this hypothesis was the same as that described in hypothesis 3. Motivation for test interpretation was determined from the Counseling Request Form administered at the end of testing. There are two classifications within this variable: (1) students who requested test interpretation, and (2) students who stated that they did not desire test interpretation.

Chi Square was used to test hypothesis 4. Students were assigned to cells on the basis of father's occupational group and motivation for test interpretation. This procedure yielded a 2 x 6 contingency table to be analyzed. This analysis was carried out only on the total sample due to the fact that when the sample is broken down by grade level cell frequencies become too small, in some cases for the use of Chi Square.

TABLE 8

RELATIONSHIP BETWEEN FATHER'S OCCUPATIONAL GROUP AND MOTIVATION

Motivation	Father's Occupational Group 1	2	3	4	5	6	Total
High	92	68	49	184	116	110	619
Low	6	5	5	27	29	12	84
Total	98	73	54	211	145	122	703

$\chi^2 = 17.784$ d.f. $= 5$ $P < .01$

Results and Discussion of Hypothesis 4.

The results of hypothesis 4 are presented in Table 8. These results indicate that the hypothesis should be rejected.

Inspection of the contingency table revealed that the relationship was such that students from the higher father's occupational groups tended to be better motivated for test interpretation.

The purpose of hypothesis 4 was essentially similar in nature to that of hypothesis 3. The purpose, in this case, was to examine the effect of socioeconomic factors as reflected by father's occupational group upon students' motivation to learn more about their ability and achievement.

It was felt that the favorable attitude toward educational achievement in the middle to upper-middle social strata would lead students from the higher father's occupational groups to be more highly motivated for test interpretation counseling. The results from the present study seem to offer support to this point of view.

Hypothesis 5.

When students are classified as to those who are in the highest one-fourth and lowest one-fourth on the basis of Accuracy scores, no significant difference exists between mean:

a. Intelligence scores of the two groups.
b. Perceived Accuracy scores of the two groups.
c. Satisfaction scores of the two groups.
d. Discrepancy scores of the two groups.
e. Congruency scores of the two groups.

Statistical Procedures for Hypothesis 5.

For hypothesis 5, criterion groups were chosen by ordering students in the total sample on the basis of Accuracy scores, and then selecting the highest 175 and lowest 175 for the criterion groups. One of the effects of this procedure was that when students were further sub-divided according to grade level, there were unequal numbers in the two groups within a particular grade level. Although criterion group comparisons typically involve equal numbers in the criterion groups, it was felt that the procedures used in the present study were desirable because they eliminated the possibility of a student being in the highest or lowest group within a grade level, but not in the total sample, and thus led to a maximum separation of the groups on the criterion variable.

Once the criterion groups were established, t-tests for difference between means were computed for each of the variables specified in the sub-hypotheses.

Results and Discussion of Hypothesis 5.

5a. Difference between Intelligence scores of the two groups.

The results shown in Table 9 indicate that the decision to reject this hypothesis is warranted at all four grade levels and Total N.

TABLE 9

t-TESTS FOR DIFFERENCE IN INTELLIGENCE SCORES OF THE HIGH- AND LOW-ACCURACY CRITERION GROUPS

	Criterion Groups					
	High		Low			
Sample	N	Mean	N	Mean	t	P
Grade 5	48	2.813	56	2.382	2.049	< .05
Grade 7	40	2.875	48	2.293	2.474	< .02
Grade 9	43	2.860	52	1.596	8.948	< .01
Grade 11	44	2.818	19	1.944	4.612	< .01
Total Sample	175	2.840	175	2.075	7.543	< .01

Apparently students who were more accurate in self-estimation had the highest Intelligence scores, whereas students who were less accurate in self-estimation had lowest Intelligence scores.

5b. Difference between Perceived Accuracy scores of the two groups.

The results shown in Table 10 indicate that this hypothesis should be retained within all grade levels, and in the Total sample. Apparently the two criterion groups are equivalent in terms of how accurate the students in them perceive their self-estimates.

5c. Difference between Satisfaction scores of the two groups.

The results which are shown in Table 11 seem to warrant the decision to reject the hypothesis within grades five and nine, but to retain it within grades seven and eleven, and in the total sample. Within grades five and nine, the students from the low-accuracy criterion group tended to report themselves as more highly satisfied with their self-estimation than were their counterparts in the high accuracy group. Within grades seven and eleven, and when the sample as a whole was analyzed, there appeared to be no difference in the satisfaction with self-estimates between the two criterion groups.

TABLE 10

t-TESTS FOR DIFFERENCE IN PERCEIVED ACCURACY SCORES
OF THE HIGH- AND LOW-ACCURACY CRITERION GROUPS

Sample	Criterion Group					
	High		Low			
	N	Mean	N	Mean	t	P
Grade 5	48	5.083	56	5.036	0.148	NS
Grade 7	40	5.075	48	4.625	1.351	NS
Grade 9	43	5.209	52	5.019	0.731	NS
Grade 11	44	5.114	19	5.474	-1.061	NS
Total Sample	175	5.120	175	4.966	0.995	NS

5d. Difference between Discrepancy scores of the two groups.

The results shown in Table 12 indicate that this hypothesis should be retained within the seventh, ninth, and eleventh grade levels, and in the Total Sample, but rejected at the fifth grade level. Apparently students at the fifth grade level from the high-accuracy criterion group had less discrepancy between "real" and "ideal" self-estimates, whereas at the other levels there was no difference in discrepancy scores between the two criterion groups.

5e. Difference between Congruency scores of the two groups.

On the basis of the results shown in Table 13, the decision to reject this hypothesis appears to be warranted. This hypothesis is rejected at all grade levels and in the total sample. In all cases, students from the

TABLE 11

t-TESTS FOR DIFFERENCE IN SATISFACTION SCORES OF THE HIGH- AND LOW-ACCURACY CRITERION GROUPS

	Criterion Group					
	High		Low			
Sample	N	Mean	N	Mean	t	P
Grade 5	48	13.938	56	15.696	-2.611	<.02
Grade 7	40	16.750	48	15.958	1.142	NS
Grade 9	43	16.907	52	18.288	-2.147	<.05
Grade 11	44	17.750	19	17.105	0.865	NS
Total Sample	175	16.269	175	16.691	-1.159	NS

TABLE 12

t-TESTS FOR DIFFERENCE IN DISCREPANCY SCORES OF THE HIGH- AND LOW-ACCURACY CRITERION GROUPS

	Criterion Group					
	High		Low			
Sample	N	Mean	N	Mean	t	P
Grade 5	48	3.771	56	5.643	-2.593	<.02
Grade 7	40	4.325	48	3.750	1.066	NS
Grade 9	43	3.674	52	3.923	-0.421	NS
Grade 11	44	3.932	19	3.474	0.716	NS
Total Sample	175	3.914	175	4.377	-1.437	NS

TABLE 13

t-TESTS FOR DIFFERENCE IN CONGRUENCY SCORES OF THE HIGH- AND LOW-ACCURACY CRITERION GROUPS

	Criterion Groups					
	High		Low			
Sample	N	Mean	N	Mean	t	P
Grade 5	48	24.146	56	27.536	-8.242	<.01
Grade 7	40	24.775	48	27.917	-7.029	<.01
Grade 9	43	24.209	52	26.808	-6.343	<.01
Grade 11	44	24.295	19	26.474	-4.104	<.01
Total Sample,	175	24.343	175	27.309	-13.482	<.01

low accuracy criterion groups showed greater change in accuracy after test interpretation counseling than did students in the high accuracy criterion groups.

The underlying purpose of hypothesis 5 was to determine some of the correlates of initial accuracy of self-estimation, and the effect of initial accuracy on outcomes of test interpretation counseling.

Froelich and Moser (1954) reported that a larger portion of brighter pupils reported their scores more accurately than did the less bright in their sample of ninth graders. This finding was upheld in the present investigation. The results of hypothesis 5a show that students in the highest quarter of Accuracy scores tended to have significantly higher intelligence scores than students in the lowest quarter. Apparently intelligence is a factor affecting accuracy of self-estimation, brighter students being more accurate in initial self-estimation.

It was originally felt by the present investigator that those students who were more unsure of their achievement and ability scores would also tend to be less accurate in their self-estimations. The results of hypothesis 5b do not support this point of view. Apparently students in the lower quarter on the basis of Accuracy scores do not perceive themselves as being more or less highly accurate than do students who were high in accuracy. Inspection of the data involved revealed that within the individual grade levels, and in the sample as a whole, the distributions of Perceived Accuracy scores were skewed in such a manner as to indicate that the majority of students tended to rate themselves as being low in accuracy of self-estimation. In seeking an explanation of this phenomenon, the self-rating scale used in the original study was inspected. It would seem that the wording of the choices on the self-rating scale was such as to "pull" for the type of distribution noted. The investigators feel that a change in wording would allow for a more "normal" distribution of Perceived Accuracy. It is suggested that in any future investigation of Perceived Accuracy of self-estimation, the scale choices used be given careful consideration as to their wording.

In terms of satisfaction with self-estimates, students in grades five and nine from the high accuracy group appeared to report themselves as less highly satisfied with self-estimates than those from the low accuracy groups. This directionality was also present in the sample as a whole, however the difference between means was not significant. It is interesting that these students from the high accuracy group would be less satisfied with their self-estimates while at the same time show less discrepancy between "real" and "ideal" self-estimates. (c.f. test for hypothesis 5d) This may be, in part, an artifact due to the fact that as these students were highly intelligent (c.f. test for hypothesis 5a) and accurate in their self-estimates, it was not likely that they would rate their "ideal" self as being lower than their "real" self. This situation creates a ceiling effect in discrepancy scores for these students.

Considering Discrepancy scores, it was originally felt that students who were high in initial accuracy of self-estimation would show less discrepancy between real and ideal self-estimates. This point of view was taken because it was felt that students who were able to accurately estimate their scores would have been more likely to have incorporated this picture of their ability and achievement into their self-image, and thus have less discrepancy between their real and ideal self-concepts. This expectancy was confirmed, at least at the fifth grade level, by the findings of hypothesis 5d. Although the difference failed to reach significance, the same directionality was preserved at the ninth grade level and in the total sample.

The relationship between initial accuracy of self-estimation and outcomes of test interpretation is clearly established by the results of hypothesis 5e. In all cases, students from the low accuracy group showed significantly higher Congruency scores than students who were high on initial accuracy. This finding is opposite to that of Gustad and Tuma (1957) who stated that client learning was positively related to initial accuracy of self-ratings. Because of the nature of the distribution of Congruency scores, (25 equals no change, greater than 25 shows increased accuracy, less than 25 shows decreased accuracy), it can also be noted that students who were low in initial accuracy showed gains after test interpretation, whereas students high in initial accuracy showed very slight losses in accuracy.

This finding would seem to be quite logical, in that students who were initially low in accuracy had great room for gains, while students high in initial accuracy had little room for gains as a result of test interpretation counseling.

As a generalized summary of the overall findings of hypothesis 5, it might be said that students who are low in initial accuracy of self-estimation are less intelligent, claim higher satisfaction with themselves, show more discrepancy between real and ideal self-estimates, and show a greater degree of change in the direction of increased accuracy as a result of test interpretation counseling than do students who are high in initial accuracy of self-estimate.

Hypothesis 6.

When students are classified as to those who are in the highest one-fourth and lowest one-fourth on the basis of Accuracy scores, no significant relationship exists between accuracy of self-estimation and:

a. Father's occupational group.
b. Father's educational level.
c. Mother's educational level.
d. Motivation for test interpretation.

Statistical Procedures for Hypothesis 6.

The accuracy rank classification used in this hypothesis was the same as that described in hypothesis 5. The father's occupational group, and father's and mother's educational level classifications were determined from the Information Sheet administered at the end of the testing. This Information Sheet and the classifications determined from it were previously described on pp. 9-10. The motivation variable used here was described under hypothesis 4.

TABLE 14

RELATIONSHIP BETWEEN ACCURACY RANK AND
FATHER'S OCCUPATIONAL GROUP

Accuracy Rank	Father's Occupation						Total
	1	2	3	4	5	6	
Highest $\frac{1}{4}$	18	17	16	52	39	33	175
Lowest $\frac{1}{4}$	26	23	12	53	30	31	175
Total	44	40	28	105	69	64	350

$\chi^2 = 4.044$ d.f. = 5 P = N.S.

Chi Square was used to test hypothesis 6. Students were assigned to the appropriate cells of a contingency table for each sub-hypothesis. Hypothesis 6a yielded a 2 x 6 table. For hypothesis 6b and 6c it was necessary to combine the lowest two and the highest two categories of educational level in order to avoid low cell frequencies. This led to these two hypotheses yielding 2 x 5 tables for analysis. Hypothesis 6d hielded a 2 x 2 table for analysis. These analyses were carried out only on the total sample due to the fact that when the sample is broken down by grade level low cell frequencies, in some cases, preclude the use of the Chi Square statistic.

Results and Discussion of Hypothesis 6.

6a. Relationship between accuracy rank and father's occupational group.

The results shown in Table 14 indicate that the decision to retain this hypothesis is warranted. Apparently there is no relationship between father's occupational group and accuracy of self-estimate expressed in high and low criterion group form.

6b. Relationship between accuracy rank and father's educational level.

The results shown in Table 15 indicate that this hypothesis should be rejected. Inspection of the cell frequencies revealed that low accuracy of self-estimation appeared to be associated with higher levels of father's education.

6c. Relationship between accuracy rank and mother's educational level.

TABLE 15

RELATIONSHIP BETWEEN ACCURACY RANK AND FATHER'S EDUCATIONAL LEVEL

Accuracy Rank	Father's Educational Level					Total
	1 & 2	3	4	5	6 & 7	
Highest $\frac{1}{4}$	47	16	96	10	6	175
Lowest $\frac{1}{4}$	45	16	68	19	27	175
Total	92	32	164	29	33	350

$\chi^2 = 20.984$ d.f. = 4 $P < .01$

TABLE 16

RELATIONSHIP BETWEEN ACCURACY RANK AND MOTHER'S EDUCATIONAL LEVEL

Accuracy Rank	Mother's Educational Level					Total
	1 & 2	3	4	5	6 & 7	
Highest $\frac{1}{4}$	27	24	94	17	13	175
Lowest $\frac{1}{4}$	29	15	92	18	21	175
Total	56	39	186	35	34	350

$\chi^2 = 4.130$ d.f. = 4 P = N.S.

TABLE 17

RELATIONSHIP BETWEEN ACCURACY RANK AND MOTIVATION FOR TEST INTERPRETATION

Accuracy Rank	Motivation		Total
	High	Low	
Highest $\frac{1}{4}$	149	26	175
Lowest $\frac{1}{4}$	163	12	175
Total	312	38	350

$\chi^2 = 5.750$ d.f. = 1 $P < .02$

The results shown in Table 16 seem to warrant the decision to retain this hypothesis. There is apparently no relationship between mother's level of education and accuracy of self-estimation expressed in high and low criterion groups.

6d. Relationship between accuracy rank and motivation for test interpretation.

The results shown in Table 17 indicate that this hypothesis should be rejected. Inspection of cell frequencies revealed that high accuracy of self-estimate tended to be associated with low motivation for test interpretation counseling.

The underlying purpose of hypothesis 6 was, as in the case of hypothesis 5, to attempt to determine some of the correlates of initial accuracy of self-estimation, and their effects upon self-estimation.

As was described in the discussion of hypothesis 5, it was originally felt that factors such as socioeconomic status as reflected by father's occupational group would show an effect on variables to self-estimation. The results of hypothesis 3a showed that in the present study there were no significant mean differences in accuracy among the various father's occupational classifications. The results from the present hypothesis lend added support to that finding. There seems to be no significant relationship between accuracy of self-estimation and father's occupational group, even when only the extremes of accuracy are considered.

In the case of father's and mother's educational levels, the results of the present investigation seem to indicate that students whose fathers tend to be in the higher educational attainment groups are less highly accurate in initial self-estimation. The variable of mother's educational level does not seem to have any effect on accuracy.

Considering the variable of motivation for test interpretation counseling, it appears to be fairly well established, in the present investigation, that students who are high in accuracy of self-estimation tend to be less well motivated for test interpretation than do students who are low in initial accuracy. This would seem to be a logical relationship. Students who are accurate in their estimates of their own ability, i.e., seem to know what their level of ability and achievement is, would feel less need to see the results of their tests than those who were not as aware of where they scored.

In generalized form, the results of hypothesis 6 would seem to indicate that students who are low in initial accuracy of self-estimation come from homes where there is a high level of educational attainment on the part of the father, and are more well motivated for test interpretation counseling. There does not, however, appear to be any relationship between initial accuracy of self-estimation and socioeconomic status as measured by father's occupational group as classified in the present investigation.

Hypothesis 7.

When students are classified as to those who are in the highest one-fourth and lowest one-fourth on the basis of congruency scores, no significant difference exists between mean:

a. Intelligence scores of the two groups.
b. Perceived accuracy scores of the two groups.
c. Satisfaction scores of the two groups.
d. Discrepancy scores of the two groups.

Statistical Procedures for Hypothesis 7.

For hypothesis 7, criterion groups were chosen by ordering students in the total sample on the basis of congruence scores, and then selecting the highest 175 and lowest 175 for the criterion groups.

Once the criterion groups were established, *t*-tests for difference between means were computed for each of the variables specified in the subhypotheses. These tests were carried out independently for each grade level, and for the total sample.

Results and Discussion of Hypothesis 7.

7a. Difference between intelligence scores of the two groups.

The results shown in Table 18 indicate that the decision to reject this hypothesis for the sample as a whole, but retain it within each individual grade level is warranted. In the total sample, students from the low congruency criterion group appear to have significantly higher intelligence scores, than those in the high congruency criterion group. This same directionality is present within each of the four grade levels, although the differences are not significant. Apparently a trend is established which reaches significance when the total sample is analyzed.

7b. Difference between Perceived accuracy scores of the two groups.

The results shown in Table 19 indicate that this hypothesis should be retained within all grade levels, and in the total sample. Apparently the two criterion groups are similar in terms of how accurate the students in them perceive their initial self-estimation.

7c. Difference between satisfaction scores of the two groups.

The results shown in Table 20 seem to warrant the decision to reject the hypothesis within grade nine, but retain it within grades five, seven, and eleven, and in the total sample. Within grade nine, the students from the high congruency criterion group tended to be initially less highly satisfied with their self-estimates than were their counterparts in the low-congruency group. Within grades five, seven, and eleven, and in the total

TABLE 18

t-TESTS FOR DIFFERENCE IN INTELLIGENCE SCORES OF THE HIGH- AND LOW-CONGRUENCY CRITERION GROUPS

	Criterion Group					
	High		Low			
Sample	N	Mean	N	Mean	*t*	P
Grade 5	47	2.489	55	2.855	-1.606	NS
Grade 7	49	2.571	36	2.722	-0.551	NS
Grade 9	50	1.980	47	2.319	-1.901	NS
Grade 11	29	2.345	37	2.611	-1.234	NS
Total Sample	175	2.343	175	2.632	-2.543	<.02

TABLE 19

t-TESTS FOR DIFFERENCE IN PERCEIVED ACCURACY SCORES OF THE HIGH- AND LOW-CONGRUENCY CRITERION GROUPS

	Criterion Group					
	High		Low			
Sample	N	Mean	N	Mean	t	P
Grade 5	47	5.085	55	4.782	0.893	NS
Grade 7	49	4.857	36	5.472	-1.709	NS
Grade 9	50	4.920	47	4.915	0.021	NS
Grade 11	29	5.103	37	5.541	-1.209	NS
Total Sample	175	4.977	175	5.120	-0.885	NS

TABLE 20

t-TESTS FOR DIFFERENCE IN SATISFACTION SCORES OF THE HIGH- AND LOW-CONGRUENCY CRITERION GROUPS

	Criterion Group					
	High		Low			
Sample	N	Mean	N	Mean	*t*	P
Grade 5	47	15.277	55	14.491	1.079	NS
Grade 7	49	16.020	36	15.333	0.874	NS
Grade 9	50	18.560	47	17.149	2.352	<.05
Grade 11	29	16.793	37	18.243	-1.816	NS
Total Sample	175	16.674	175	16.171	1.297	NS

sample, there appeared to be no difference in initial satisfaction between the two criterion groups.

7d. Difference between Discrepancy scores of the two groups.

The results shown in Table 21 indicate that this hypothesis should be retained within all grade levels and in the total sample. Apparently there is no difference between the two criterion groups in terms of initial discrepancy between "real" and "ideal" self-estimates.

The underlying purpose of hypothesis 7 was to determine the effects of certain variables related to the process of self-estimation and test-interpretation counseling on outcomes of such counseling as reflected by gains in accuracy of self-estimates.

TABLE 21

t-TESTS FOR DIFFERENCE IN DISCREPANCY SCORES OF THE HIGH- AND LOW-CONGRUENCY CRITERION GROUPS

	Criterion Groups					
	High		Low			
Sample	N	Mean	N	Mean	t	P
Grade 5	47	5.489	55	4.236	1.847	NS
Grade 7	49	4.061	36	3.944	0.180	NS
Grade 9	50	4.300	47	3.957	0.635	NS
Grade 11	29	3.793	37	4.135	-0.556	NS
Total Sample	175	4.469	175	4.080	1.232	NS

No clear-cut pattern of effects seems to emerge regarding the variables studied. Apparently students from the low congruency criterion group had higher intelligence scores than those in the high congruency group. This would appear to be a logical extension of the finding that students who were more intelligent tended to be more accurate initially (c.f. test for hypothesis 5a), and that those who were more accurate initially showed less gain in accuracy; i.e., lower congruency scores. (c.f. test for hypothesis 5e).

There also appeared to be an indication that students, within the ninth grade level, from the high congruency criterion group were initially less satisfied with their self-estimates than those in the low congruency group. As this finding is not reflected in any of the other grade levels, or in the sample as a whole, it would seem likely that it is due more to the operation of chance factors rather than consistent variation among students in the criterion groups across all grade levels.

Hypothesis 8.

When students are classified as to those who are in the highest one-fourth and lowest one-fourth on the basis of congruency scores, no significant relationship exists between congruency and:

a. Father's occupational group.
b. Father's educational level.
c. Mother's educational level.
d. Motivation for test interpretation.

Statistical Procedures for Hypothesis 8.

The congruency rank classification used in this hypothesis was the same as that described in hypothesis 7. The father's occupational group, father's and mother's educational levels, and motivation classifications were those described on pp.11-12.

Chi Square was used to test hypothesis 8. Students were assigned to the appropriate cells of a contingency table for each sub-hypothesis. Hypothesis 8a yielded a 2 x 6 table. For hypotheses 8b and 8c it was necessary to combine the lowest two and the highest two categories respectively of educational level in order to avoid low cell frequencies. This led to those two hypotheses yielding 2 x 5 tables for analysis. Hypothesis 8d yielded a 2 x 2 table for analysis. These analyses were carried out only on the total sample due to the fact that when the sample is broken down by grade level low cell frequencies, in some cases, preclude the use of the Chi Square statistic.

Results and Discussion of Hypothesis 8.

8a. Relationship between congruency rank and father's occupational group.

The results shown in Table 22 indicate that the decision to retain this hypothesis is warranted. Apparently there is no relationship between father's occupational group and gain in accuracy after test interpretation.

8b. Relationship between congruency rank and father's educational level.

The results shown in Table 23 indicate that this hypothesis should be rejected. Inspection of the cell frequencies revealed that high congruency tended to be associated with high levels of father's education.

8c. Relationship between congruency rank and mother's educational level.

The results shown in Table 24 seem to warrant the decision to retain this hypothesis. Apparently there is no significant tendency for association between mother's educational level and gain in accuracy after test interpretation.

8d. Relationship between congruency rank and motivation for test interpretation.

TABLE 22

RELATIONSHIP BETWEEN CONGRUENCY RANK AND FATHER'S OCCUPATIONAL GROUP

Congruency Rank	Father's Occupational Group						Total
	1	2	3	4	5	6	
Highest $\frac{1}{4}$	25	19	13	50	38	30	175
Lowest $\frac{1}{4}$	19	21	6	60	39	30	175
Total	44	40	19	110	77	60	350

$\chi^2 = 4.286$ d.f. = 5 P = N.S.

TABLE 23

RELATIONSHIP BETWEEN CONGRUENCY RANK AND FATHER'S EDUCATIONAL LEVEL

Congruency Rank	Father's Educational Level					Total
	1 & 2	3	4	5	6 & 7	
Highest $\frac{1}{4}$	48	18	64	20	25	175
Lowest $\frac{1}{4}$	47	11	90	13	14	175
Total	95	29	154	33	39	350

$\chi^2 = 10.674$ d.f. = 4 $P < .05$

TABLE 24

RELATIONSHIP BETWEEN CONGRUENCY RANK AND MOTHER'S EDUCATIONAL LEVEL

Congruency Rank	Mother's Educational Level					Total
	1 & 2	3	4	5	6 & 7	
Highest $\frac{1}{4}$	29	26	83	19	18	175
Lowest $\frac{1}{4}$	24	17	100	18	16	175
Total	53	43	183	37	34	350

$\chi^2 = 4.080$ d.f. = 4 P = N.S.

The results shown in Table 25 indicate that this hypothesis should be rejected. Inspection of cell frequencies revealed that high congruency tended to be associated with high motivation for test interpretation.

TABLE 25

RELATIONSHIP BETWEEN CONGRUENCY RANK AND
MOTIVATION FOR TEST INTERPRETATION

Congruency Rank	Motivation		Total
	High	Low	
Highest $\frac{1}{4}$	163	12	175
Lowest $\frac{1}{4}$	150	25	175
Total	313	37	350

$\chi^2 = 5.073$ d.f. = 1 $P < .01$

The purpose of hypothesis 8 was to determine the effect of certain parental socioeconomic factors on outcomes of test interpretation counseling as expressed by changes in accuracy after counseling. Also studied was the effect of motivation for test interpretation on such outcomes.

Two of the variables, father's educational level and motivation for test interpretation, seem to be related to those outcomes. Apparently students whose fathers have higher levels of educational attainment make greater changes in the direction of increased accuracy of self-estimate following test interpretation. This finding is logically related to other findings in the present study in that students from homes where fathers are in the high levels of educational attainment are less accurate in initial self-estimations (hypothesis 6b), and students who are low in initial accuracy make greater changes following test interpretation. (hypothesis 5e)

The finding regarding motivation, i.e., higher motivation is associated with higher congruency, would also seem to fit this pattern, in that high motivation is associated with low initial accuracy (hypothesis 6d), which is, in turn, associated with greater congruency as was stated above.

In summary, the only apparent pattern emerging from the results of this hypothesis is that students who make greater gains in accuracy following test interpretation; (1) come from homes where the father has a higher level of educational attainment, and (2) are more highly motivated for test interpretation counseling.

Both of these findings seem to lend support to the findings of the tests for hypotheses 5 and 6, and the finding regarding motivation lends support to the commonly held belief that test interpretation counseling is more effective with students who request it than with students who have no particular desire for it.

Hypothesis 9.

When students are classified as to those who are accurate-estimators, under-estimators, or over-estimators, no significant differences exist among mean:

a. Intelligence scores of the groups.
b. Accuracy scores of the groups.
c. Perceived accuracy scores of the groups.
d. Satisfaction scores of the groups.
e. Discrepancy scores of the groups.
f. Congruency scores of the groups.

Statistical Procedures for Hypothesis 9.

To test hypothesis 9, students were classified into groups according to their consistency of direction of self-estimate. Those who over-estimated their scores on four or more of the six subtests were classified

TABLE 26

ANALYSIS OF VARIANCE OF INTELLIGENCE SCORES OF STUDENTS
FROM THREE ESTIMATION RANKS

Sample	Source of Variation	Sum of Squares	df	Mean Square	F	P
Grade 5	Between	69.062	2	34.531	57.647	<.01
	Within	111.521	186	0.599		
	Total	180.583	188			
Grade 7	Between	93.692	2	46.846	68.488	<.01
	Within	118.490	173	0.684		
	Total	212.182	175			
Grade 9	Between	57.389	2	28.694	66.730	<.01
	Within	86.548	201	0.430		
	Total	143.937	203			
Grade 11	Between	24.729	2	12.364	22.118	<.01
	Within	72.203	129	0.559		
	Total	96.932	131			
Total Sample	Between	255.089	2	127.544	233.369	<.01
	Within	399.083	698	0.571		
	Total	654.172	700			

TABLE 27

ANALYSIS OF VARIANCE OF ACCURACY SCORES OF STUDENTS FROM THREE ESTIMATION RANKS

Sample	Source of Variation	Sum of Squares	df	Mean Square	F	P
Grade 5	Between	740.611	2	370.305	61.768	<.01
	Within	1115.199	186	5.995		
	Total	1855.810	188			
Grade 7	Between	595.343	2	297.671	68.889	<.01
	Within	747.567	173	4.321		
	Total	1342.910	175			
Grade 9	Between	654.302	2	327.151	86.069	<.01
	Within	764.105	201	3.801		
	Total	1418.407	203			
Grade 11	Between	314.368	2	157.184	39.022	<.01
	Within	523.723	130	4.028		
	Total	838.091	132			
Total Sample	Between	2379.370	2	1189.685	255.571	<.01
	Within	3258.648	700	4.655		
	Total	5638.018	702			

as over-estimators, those who under-estimated themselves on four or more scales were classified as under-estimators, and the remaining students were classified as accurate-estimators. Using these classifications as the treatment groups, a single classification analysis of variance was computed for each of the variables specified in the sub-hypotheses. These analyses were carried out separately for each grade level, and for the total sample. In those cases where significant F ratios were obtained, Kramer's extension of Duncan's new multiple range test (Kramer, 1956) was applied to make multiple comparisons among the group means.

Results and Discussion of Hypothesis 9.

9a. Differences among intelligence scores of the groups.

The results shown in Table 26 indicate that the decision to reject this hypothesis at all grade levels and in the total sample is warranted. The multiple range test revealed that in the fifth and seventh grades and in the total sample, the means of the three groups were significantly different for all comparisons. In the ninth and eleventh grade samples, the under-estimators scored significantly higher than either the accurate or over-estimators, but the difference between the latter two groups was not significant. In all cases the under-estimator group had the highest mean intelligence score and the over-estimator group the lowest, with the accurate-estimators mean lying between the two. The unusually high F ratios obtained in this analysis may be an artifact due to the fact that highly intelligent students have more scope for underestimating, and low intelligence students have more scope for overestimation.

9b. Differences among accuracy scores of the groups.

The results shown in Table 27 indicate that this hypothesis should be rejected within any specific grade level and in the sample as a whole. This finding is not surprising in that the method of classification of groups was partly dependent on accuracy scores. The results of the multiple range test indicated that accurate estimators consistently scored significantly higher on accuracy than either under or over-estimators, which is to be expected. There was a general tendency for under-estimators to score higher than over-estimators in accuracy of self-estimate. This difference was not significant within a specific grade level, but was when the total sample was analyzed.

9c. Differences among perceived accuracy scores of the groups.

The results shown in Table 28 seem to warrant the decision to reject this hypothesis within grade seven, but retain it within the remaining three grade levels and in the total sample. Application of the multiple range test to the analysis for the seventh grade revealed that the only significant comparison was between the accurate and under-estimating groups with the under-estimators having the higher scores.

9d. Differences among satisfaction scores of the groups.

TABLE 28

ANALYSIS OF VARIANCE OF PERCEIVED ACCURACY SCORES OF STUDENTS FROM THREE ESTIMATION RANKS

Sample	Source of Variation	Sum of Squares	df	Mean Squares	F	P
Grade 5	Between	0.512	2	0.256	0.092	NS
	Within	515.199	187	2.755		
	Total	515.711	189			
Grade 7	Between	15.952	2	7.976	3.108	<.05
	Within	443.958	173	2.566		
	Total	459.910	175			
Grade 9	Between	4.042	2	2.021	1.244	NS
	Within	326.561	201	1.624		
	Total	330.603	203			
Grade 11	Between	2.221	2	1.110	0.576	NS
	Within	250.306	130	1.925		
	Total	252.527	132			
Total Sample	Between	7.472	2	3.736	1.681	NS
	Within	1555.709	700	2.222		
	Total	1563.181				

The results shown in Table 29 appear to warrant the decision to retain this hypothesis within the seventh and eleventh grade levels, but reject it within the fifth and ninth grade levels and in the sample as a whole. The multiple range tests revealed that within the fifth grade the under-estimators scored significantly lower than either the accurate or over-estimators, however the difference between the latter two groups was not significant. In the ninth grade the only significant difference noted was that under-estimators scored significantly lower than accurate estimators. In the total sample, the means of the three groups were significantly different for all comparisons, with the over-estimators scoring highest and the under-estimators lowest.

9e. Differences among discrepancy scores of the groups.

On the basis of the results shown in Table 30, the decision to retain this hypothesis appears to be warranted. This hypothesis is retained at all grade levels, and in the sample as a whole. Apparently students from the three estimation ranks are homogeneous with respect to discrepancy between "real" and "ideal" self-ratings.

TABLE 29

ANALYSIS OF VARIANCE OF SATISFACTION SCORES OF STUDENTS FROM THREE ESTIMATION RANKS

Sample	Source of Variation	Sum of Squares	df	Mean Square	F	P
Grade 5	Between	132.646	2	66.323	4.673	<.05
	Within	2654.033	187	14.192		
	Total	2786.679	189			
Grade 7	Between	83.601	2	41.800	2.914	NS
	Within	2481.127	173	14.341		
	Total	2564.728	175			
Grade 9	Between	81.717	2	40.858	4.246	<.05
	Within	1933.867	201	9.621		
	Total	2015.584	203			
Grade 11	Between	18.457	2	9.228	0.913	NS
	Within	1312.792	130	10.098		
	Total	1331.249	132			
Total Sample	Between	412.758	2	206.379	15.466	<.01
	Within	9341.072	700	13.344		
	Total	9753.830				

TABLE 30

ANALYSIS OF VARIANCE OF DISCREPANCY SCORES OF STUDENTS FROM THREE ESTIMATION RANKS

Sample	Source of Variation	Sum of Squares	df	Mean Square	F	P
Grade 5	Between	3.538	2	1.769	0.142	NS
	Within	2320.615	187	12.409		
	Total	2324.153	189			
Grade 7	Between	21.332	2	10.666	1.389	NS
	Within	1328.117	173	7.676		
	Total	1349.449	175			
Grade 9	Between	0.296	2	0.148	0.002	NS
	Within	1349.704	201	6.714		
	Total	1350.000	203			
Grade 11	Between	23.413	2	11.706	2.119	NS
	Within	718.106	130	5.523		
	Total	741.519	132			
Total Sample	Between	8.738	2	4.369	0.525	NS
	Within	5824.793	700	8.321		
	Total	5833.531	702			

9f. *Differences among congruency scores of the groups.*

The results shown in Table 31 indicate that this hypothesis should be rejected at all grade levels and in the total sample. Application of the multiple range test revealed that accurate estimators make significantly less gain in accuracy following test interpretation than either under or over-estimators, but the difference between the latter two groups is not significant.

TABLE 31

ANALYSIS OF VARIANCE OF CONGRUENCY SCORES OF
STUDENTS FROM THREE ESTIMATION RANKS

Sample	Source of Variation	Sum of Squares	df	Mean Square	F	P
Grade 5	Between	185.087	2	92.543	20.357	$<.01$
	Within	850.287	187	4.546		
	Total	1035.374	189			
Grade 7	Between	129.378	2	64.689	12.686	$<.01$
	Within	882.259	173	5.099		
	Total	1011.637	175			
Grade 9	Between	140.501	2	70.250	17.013	$<.01$
	Within	830.088	201	4.129		
	Total	970.589	203			
Grade 11	Between	48.961	2	24.490	7.507	$<.01$
	Within	424.177	130	3.362		
	Total	473.138	132			
Total Sample	Between	448.844	2	224.422	51.331	$<.01$
	Within	3060.952	700	4.372		
	Total	3509.796	702			

Hypothesis 9 was essentially a descriptive hypothesis. Students can be inaccurate in their self-estimations in two ways: they can over-estimate their abilities, or they can under-estimate them. It was felt by the present investigators that there may be systematic differences among students who characteristically fall in these groups. Therefore, in addition to comparing students who were highly accurate with those who were inaccurate (hypotheses 5 and 6), it was deemed necessary to compare students who were accurate in self-estimations with students who were inaccurate in both directions, *i.e.* make a three-way comparison among accurate-estimators, under-estimators, and over-estimators. .

The results of testing this hypothesis led to the following descriptions of these groups: Over-estimators appear to be lower in terms of intelligence, have lower initial accuracy of self-estimation, are the most highly satisfied with their self-estimates, and show high gain in accuracy following test interpretation. Accurate-estimators are the highest in initial accuracy of self-estimation, the lowest in gains in accuracy following test interpretation, and the middle group on intelligence and satisfaction scores. Under-estimators have the highest intelligence scores, are least satisfied with their self-ratings, are the middle group on initial accuracy, and show some gain in accuracy following test interpretation.

Of these descriptions, that of the over-estimating group was clearly supported by the results of testing previous hypotheses in the present study. The variables in this group are logically related in terms of the findings of hypothesis 5a (low intelligence-low accuracy), 5c (low accuracy-high satisfaction), 5e (low accuracy-high congruency), and 7a (low intelligence-high congruency). The descriptions of the other two groups, accurate and under-estimators, are supported by hypothesis 5c and 7a. It is possible to find additional support for them, however, by logical extension of the majority of the hypotheses supporting the description of over-estimators.

Hypothesis 10.

When students are classified as to those who are accurate-estimators, under-estimators, or over-estimators, no significant relationship exists between estimation rank and:

a. Father's occupational group.
b. Father's educational level.
c. Mother's educational level,
d. Motivation for test interpretation.

Statistical Procedures for Hypothesis 10.

The estimation classification system used in this hypothesis was the same as that described in hypothesis 9. The father's occupational group, father's and mother's educational levels, and motivation classifications are those described on pp. 9-11.

Chi Square was used to test hypothesis 10. Students were assigned to the appropriate cells of a contingency table for each variable specified in the sub-hypotheses. For hypothesis 10a and 10b it was necessary to combine the highest two and the lowest two categories, respectively, in order to avoid low cell frequencies. This procedure yielded 3 x 5 tables for analysis. In hypothesis 10c it was necessary to combine the lowest two and highest three classifications, respectively, which yielded a 3 x 4 table for analysis. Hypothesis 10d yielded a 3 x 2 table for analysis.

These Chi Square analyses were carried out only on the sample as a whole due to the fact that when the sample is broken down by grade level, low cell frequencies preclude the use of the Chi Square statistic.

2

TABLE 32

RELATIONSHIP BETWEEN ESTIMATION RANK AND FATHER'S OCCUPATIONAL GROUP

Estimation Rank	Father's Occupational Group					
	1 & 2	3	4	5	6	Total
Over	7	6	10	14	10	47
Accurate	101	34	122	86	81	424
Under	63	14	79	45	31	232
Total	171	54	211	145	122	703

$\chi^2 = 13.166$ d.f. = 8 P = N.S.

TABLE 33

RELATIONSHIP BETWEEN ESTIMATION RANK AND FATHER'S EDUCATIONAL LEVEL

Estimation Rank	Father's Educational Level					
	1 & 2	3	4	5	6 & 7	Total
Over	15	9	13	5	5	47
Accurate	106	36	206	33	43	424
Under	50	15	111	21	35	232
Total	171	60	330	59	83	703

$\chi^2 = 17.298$ d.f. = 8 $P < .05$

TABLE 34

RELATIONSHIP BETWEEN ESTIMATION RANK AND MOTHER'S EDUCATIONAL LEVEL

Estimation Rank	Mother's Educational Level				
	1 & 2	3	4	5, 6, & 7	Total
Over	10	12	19	6	47
Accurate	64	43	230	87	424
Under	31	13	131	57	232
Total	105	68	380	150	703

$\chi^2 = 23.132$ d.f. = 6 $P < .01$

Results and Discussion of Hypothesis 10.

10a. Relationship between estimation rank and father's occupational group.

The results shown in Table 32 indicate that the decision to retain this hypothesis is warranted. Apparently there is no relationship between father's occupational group and student's tendency to over or underestimate test scores.

10b. Relationship between estimation rank and father's educational level.

The results shown in Table 33 indicate that this hypothesis should be rejected. Inspection of cell frequencies revealed that there was a tendency for under-estimators to have fathers from higher educational levels, and the opposite to be true for over-estimators.

10c. Relationship between estimation rank and mother's educational level.

The results shown in Table 34 seem to warrant the decision to reject this hypothesis. Inspection of cell frequencies indicates that overestimation tends to be associated with lower mother's education, and underestimation with higher levels of mother's education.

10d. Relationship between estimation rank and motivation for test interpretation.

The results shown in Table 35 indicate that this hypothesis should be rejected. Inspection of cell frequencies reveals that all three groups appear to be well motivated for test interpretation counseling, with the over-estimators tending to be associated with the highest motivation.

TABLE 35

RELATIONSHIP BETWEEN ESTIMATION RANK AND
MOTIVATION FOR TEST INTERPRETATION

Estimation Rank	Motivation		Total
	High	Low	
Over	40	7	47
Accurate	360	64	424
Under	221	11	232
Total	621	82	703

$\chi^2 = 16.021$ d.f. = 2 $P < .001$

The purpose of hypothesis 10 was, as in the case of hypothesis 9, to attempt to describe the characteristics of over and under-estimators when they are compared along with accurate-estimators on the basis of parental socioeconomic factors and motivation for test interpretation counseling.

The results of testing this hypothesis lead to the following descriptions of these groups: Over-estimators appear to be very well motivated for test interpretation, and to have fathers and mothers from the lower educational levels. This placement of students with parents from the lower educational levels in the over-estimation group who are also low in initial accuracy of self-estimate (hypothesis 9), appears to be in contradiction to hypothesis 6b which indicated that students low in initial accuracy had parents from high levels of education. This contradiction may be due, in part, to the fact that in hypothesis 6 only the high and low criterion groups of accuracy were used, while in hypothesis 9 students from all levels of accuracy were included. Apparently, when the students from the middle ranges of accuracy are included, and analysis is completed based on a third classification (estimation rank), the pattern of association is shifted enough to warrant differing conclusions.

Under-estimators apparently are well motivated for test interpretation and have fathers and mothers who are in the higher levels of educational attainment.

Conclusions

The procedures reported in this chapter were designed to test hypotheses that were formulated to investigate the inter-relationships of certain variables assumed to be related to the process of self-estimation, and the relationship of these variables to outcomes of test interpretation counseling measured by changes in self-estimation. Within the limitations of this study, the following conclusions seem to be warranted regarding the variables investigated:

1. Students are adequately able to distinguish between grades earned in school subjects and performance on standardized achievement tests.

2. Students who are initially highly accurate in self-estimation tend to have higher intelligence, are less satisfied with their self-estimations, have less discrepancy between "real" and "ideal" self-estimates, are well motivated for test interpretation counseling, and make the least gain in accuracy following test interpretation counseling which would be expected. Because of the way in which the variable of initial accuracy was tested, i.e. high Vs. low criterion groups, the opposite could be concluded regarding students who are initially highly inaccurate in self-estimation.

3. In terms of gain in accuracy after test interpretation counseling, students with higher congruency scores seem to be lower in

intelligence, have fathers in the higher levels of educational attainment, and are higher in motivation for test interpretation counseling. Because of the method of analysis, the opposite could be concluded regarding students with low congruency scores.

4. Father's occupational group, as classified in the present investigation, appeared to play no significant role in terms of relationship to the other variables investigated except that students with fathers in the higher occupational classifications are more highly motivated for test interpretation counseling, and have higher perceived accuracy of self-estimation.

5. Perhaps the most interesting conclusions to be drawn from the present study are the descriptions of under-estimating and over-estimating students.

   Under-estimators seem to be high ability students who come from home environments where there is a relatively high level of educational attainment on the part of their fathers and mothers. These students are moderately accurate in their initial self-estimations, are the least satisfied with their self-estimates, are well motivated for test interpretation counseling, and achieve moderate gains in accuracy following this counseling.

   Over-estimators, on the other hand, appear to be low ability students who come from homes where there is a relatively low level of educational attainment on the part of their parents. These students are highly inaccurate in their initial self-estimations, but claim a higher degree of satisfaction with these self-ratings. However, they seem to be very well motivated for test interpretation counseling, and make the greatest gains in accuracy of self-estimate as a result of such counseling.

6. It is also interesting to note that there appears to be a factor operating in self-estimation that is an instance of the phenomenon of regression toward the mean. High ability students tend to underestimate themselves, whereas, low ability students tend to overestimate themselves.

Since the present study was designed specifically to investigate the relationships between selected variables, accuracy of self-estimate and changes in accuracy of self-estimate during the test interpretation process, it is difficult to draw implications concerning specific counselor practice. However, from the findings of the present investigation it would seem to be apparent that there are very definite differences among students in terms of their self-understanding regarding ability and achievement. In addition it appears that there are differences not only in terms of degree of accuracy of self-estimation, but also differences between students who are inaccurate in opposite fashions, i.e. under and over-estimating students. Also, students from differing home backgrounds seem to approach testing and test interpretation in different ways.

These differences manifest themselves in students' attitudes toward test interpretation and in the gains in self-understandings that students make as a result of careful test interpretation.

Counselors using test results with their students in an attempt to help them become aware of their strengths and weaknesses so they may make realistic educational and vocational choices should encourage students to estimate their scores and to look for specific discrepancies in estimates, and look beyond the test results and students' grades in performing test interpretation counseling. They should note not only how accurate a student is in estimating his test results, but also the direction of his error, and discuss these discrepancies with the student.

In summary, the fact that a counselor tells his counselee the test results, and explains what they mean in terms of percentile ranks, is no guarantee that the student will alter his self-image or make meaningful use of the information. Counselors should expect students to be unique in their approach to test interpretation, and the uses they make of the information given them. Apparently some assimilate this information quite readily, while others do not alter their self-image to any appreciable degree. It is important for counselors to understand the student and his attitudes toward testing and test-interpretation in order to perform adequately in test interpretation counseling.

Counselors might do well to spend less time working with those students who are initially quite accurate in self-estimation and devote more time to working with relatively low ability students from home environments where there is a low level of educational attainment on the part of the parents. From the wealth of recent literature on early school leavers, one is led to believe that these kinds of students are the ones who do not have an accurate image of themselves, and perhaps this may account in part for the fact that school work is not very meaningful and eventually they drop out of school. Findings from the present study however, indicate that these students report themselves as highly motivated for test interpretation counseling, and make changes in self-estimate based on this counseling. Perhaps if they received more attention from counselors, and were guided into educational programs more suited to their needs, many who might otherwise drop out would not become early school leavers.

Limitations of the Study

A limitation apparent in the present investigation, and one which is characteristic of many investigations of counseling and test interpretation, is the use of a self-rating device. To the extent that one can assume that students are basically accurate in their self-reports this approach should yield information which is usable in analyzing self-evaluation and changes in it due to counseling. To the extent that students do not take the task of filling out a self-report form seriously, and merely mark the instrument in a cavalier or thoughtless fashion, self-rating data can be misleading or useless in making any analysis of changes in self-image. There is no reason to believe that students in the present investigation were dishonest in their self-evaluations; however, on the section of the scales dealing with students' perceived accuracy of self-estimate, students' scores were skewed in such a manner as to indicate that as a group they tended to rate themselves as being low in accuracy of self-estimation. This phenomenon appeared to be due, in part, to the wording of the scales which seems to "pull" for this type of a distribution. Perhaps a change in wording would allow for a more normal distribution of perceived accuracy.

The father's occupational group classification system used in the present study was designed to attempt an indirect classification of students along a socioeconomic dimension. As no systematic relationships were found between this variable and the other variables studied, it would seem that this was not a valid or useful classification system. There is a possibility that real differences between students from various levels of this dimension were masked because of the relatively broad areas of classification.

The "outcome" variable used in the present investigation was change in accuracy from before to after test interpretation counseling. Although distinct relationships were found between this and the other variables studied, it is not a sufficient outcome measure. Behavior changes in students also should be considered in assessing outcomes of test interpretation.

Suggestions for Further Research

The present investigation was designed to assess the presence or absence of relationships among variables presumed to be important in the test interpretation process. Useful information might be gathered if further research along these lines would attempt to evaluate further cause-effect relationships such as why some students are more accurate in self-estimation than others, or why some students consistently underestimate themselves, while others overestimate their achievement and ability.

If the variable of socioeconomic status is to be included in any future study, it is recommended that a more accurate classification system be used than was the case in the present investigation.

Even though a student is more accurate in estimating his test scores after test interpretation, he may not actually make use of the information he has gained. For this reason it would be quite important for future studies of the test interpretation process to develop a criterion for "outcome" based on changes in the attitudes and/or behaviors of students who gained in accuracy of self-estimation as a result of test interpretation counseling. It would be productive to know not only which students gained in accuracy, but also what differences there were between those who actually used this information, and those who did not.

REFERENCES

Adamek, E. G. The effects of testing and test interpretation on selected self-perceptions. Unpublished doctoral thesis, University of Illinois, 1961.

Belovsky, D., McMasters, W., Shorr, J., and Singer, S. L. Individual and group counseling. *Personnel and Guidance Journal*, 1953, 31:363-365.

Biersdorf, Kathryn R. The effectiveness of two group vocational guidance treatments. Doctoral thesis, University of Maryland, 1958, *Dissertation Abstracts*, 1958, 19 (1):162-164.

Bonsall, Marcella, Meyers, C. E., and Thorpe, L. P. *What I Like To Do: An Inventory of Children's Interests*, Examiner Manual. Chicago: Science Research Associates, Inc., 1954.

Dressel, Paul, and Matteson, R. W. The effects of client participation in test interpretation. *Educational and Psychological Measurement*, 1950, 10:693-706.

Edwards, A. L. Experimental Design in Psychological Research, New York: Rinehart, 1960.

Froelich, C. P., and Moser, W. E. Do counselees remember test scores? *Journal of Counseling Psychology*, 1954, 1:149-152.

Gustad, J. W., and Tuma, A. H. The effects of different methods of test interpretation on client-learning in counseling. *Journal off Counseling Psychology*, 1957, 4:313-317.

Herrick, V. E. In *The Fifth Mental Measurements Yearbook*, O. K. Buros, Ed. Highland Park: The Gryphon Press, 1959, 30-34.

Holmes, June E. The comparison of four techniques used in presenting test information to freshmen students. *Dissertation Abstracts*, 1961, 21 (11): 3379.

Johnson, D. G. Effects of vocational counseling on self-knowledge. *Educational and Psychological Measurement*, 1953, 13:330-338.

Kamm, R. B., and Wrenn, C. G. Client acceptance of self-information in counseling. *Educational and Psychological Measurement*, 1950, 10:32-34.

Kramer, C. Y. Extension of multiple range tests to group means with unequal numbers of replications. *Biometrics*, 1956, 12:307-310.

Lallas, J. E. A comparison of three methods of interpretation of results of achievement tests to pupils. *Dissertation Abstracts*, 1956, 16 (10): 1842.

Lane, D. A comparison of two techniques of interpreting test results to clients in vocational counseling. *Dissertation Abstracts*, 1952, 12:591-592.

LeFever, D. W. In The Fifth Mental Measurements Yearbook, O. K. Buros, Ed. Highland Park: The Gryphon Press, 1959, 36-37.

Lister, J. L., and Ohlsen, M. M. The effects of orientation to testing on motivation for and outcomes of test interpretation, Cooperative Research Project No. 1344, College of Education, University of Illinois, 1962.

Remmer, H. H. In The Fifth Mental Measurements Yearbook, O. K. Buros, Ed. Highland Park: The Gryphon Press, 1959, 36-37.

Rothney, J. W. M. In The Fifth Mental Measurements Yearbook, O. K. Buros, Ed. Highland Park: The Gryphon Press, 1959, 122.

Saper, Bernard. The interpretation of tests in counseling students. Educational Record, 1961, 42:117-121.

Singer, S. L., and Stefflre, B. Analysis of the self-estimate in the evaluation of counseling. Journal of Counseling Psychology, 1954. 1:252-255.

Torrence, E. P. Practical use of knowledge about self-concept. Educational and Psychological Measurement, 1954, 14:120-127.

Tuma, A. H., and Gustad, J. W. The effects of client and counselor personality on client learning. Journal of Counseling Psychology, 1957, 4:136-143.

Walker, Helen M., and Lev, Joseph. Statistical Inference. New York, Holt, Rinehart and Winston: 1953.

Wright, E. W. A comparison of individual and multiple counseling for test interpretation interviews. Unpublished doctoral thesis, University of California, Berkeley, 1957.

FACTORS ASSOCIATED WITH TEACHER OVER- OR UNDERESTIMATION OF STUDENT INTELLECTUAL ABILITY

Richard E. Pearson and Merle M. Ohlsen

Purpose of the Study

This study was undertaken in the belief that the informal judgments which teachers make of individual students' intellectual ability, and the accuracy of these judgments, exert a significant influence upon the character of the teachers' day-to-day classroom interaction with the students. As Gage (1955) has stated:

> Conceivably, teachers who make such judgments more accurately will, other things being held equal, teach more effectively. Their assignments, their explanations, and their discussions will be more appropriate to their pupils' abilities.

Assuming that the accuracy or inaccuracy of such judgments has important consequences, it is believed that the knowledge of factors associated with teacher inaccuracy is useful, important knowledge, since, hopefully, this knowledge may be used by teacher educators and supervisors (as well as by teachers themselves) to help alert teachers to various biases which may result in such errors of judgment.

The present study explores relationships between teachers' perceptions (as reflected in their ratings) of students and the accuracy of their estimates. The study attempts to determine if there exist any systematic relationships between the level at which teachers rate students in the areas of personal appearance, classroom conduct, social and economic background, and academic achievement, and the accuracy with which they estimate the same students' intellectual ability.

While it is realized that these four areas do not exhaust those which might shed light upon factors associated with the accuracy of teachers' estimates of students' intellectual ability, it is the opinion of the present investigators that an exploration of the areas selected can provide information useful to an understanding of attitudes which teachers hold toward their students: attitudes which often determine the level at which teachers expect individual students to perform academically, and influence the type of educational experience to which they will expose the student.

Specifically, this exploration will be carried out by means of the following hypotheses:

Hypothesis 1. There will be a significant, positive relationship between discrepancy scores representing the degree and direction (i.e. under- or overestimation) of discrepancy between teacher estimates of students' intellectual ability and the students' actual performance on the Henmon-Nelson and teacher ratings in each of the following areas: 1) personal appearance, 2) classroom conduct, 3) social and economic background, and 4) academic achievement.

Hypothesis 2. There will be a significant, systematic differences between the mean discrepancy scores (scores which represent the degree and direction of discrepancy between a teacher's estimate of a student's intellectual ability and the student's actual performance on the Henmon-Nelson), of students rated high (i.e. /1 s.d.) and those rated low (i.e. -1 s.d.) in each of the four areas mentioned in Hypothesis 1. More specifically:

a. the discrepancy scores of students rated high in the area of personal appearance will be significantly higher (and will indicate over-estimation) than those of students rated low (whose discrepancy scores will indicate under-estimation),
b. the discrepancy scores of students rated high in the area of classroom conduct will be significantly higher (and will indicate over-estimation) than those of students rated low (whose disdrepancy scores will indicate under-estimation),
c. the discrepancy scores of students rated high in the area of social and economic background will be significantly higher (and will indicate overestimation) than those of students rated low (whose discrepancy scores will indicate under-estimation),
d. the discrepancy scores of students rated high in the area of academic achievement will be significantly higher (and will indicate over-estimation) than those of students rated low (whose discrepancy scores will indicate under-estimation).

Related Research

The literature relating to the measurement of intelligence contains many studies which explore the relationship between the judgments which teachers make of students' intellectual ability and the students' actual performance on a criterion measure, (see below*). Generally, these studies have found positive relationship between these judgments and actual IQ scores. Correlation coefficients (both Pearsonian and rank order) indicating the degree of relationship between the two variables have generally ranged from about .40 to .75.

---

*(Ackerson and Estabrooks, 1928; Wilson, 1925; Schwarz, 1934: Olander and Walker, 1936; Hubbard and Flescher, 1953; Moore, 1937; Jordan, 1923.)

The literature also contains a number of studies which have gone beyond the exploration of the relationship between teacher judgments and students' actual test scores, to examine factors related to under- and/or over-estimation. Gibbons (1938), for example, found that teachers were least accurate in estimating the intelligence of students not in the average ability range, tending to overestimate the ability of the "dull" group and to underestimate the ability of the superior group. Lewis (1947) obtained similar findings with students designated as "retarded" by their teachers. Hanna (1950), who made estimates of clients' intellectual ability within the context of a vocational guidance interview found that underestimation of IQ was usually associated with a notation by the counselor that the client was taciturn and quiet. Also, counselor notation of pleasing appearance and personality traits, youth and good conversational ability in the client tended to be associated with overestimation of his IQ.

Procedures

The 562 students who made up the student sample in this study are a portion of the sample used by Lister and Ohlsen (1962) in their study of factors influencing the outcomes of test interpretation. The main group of students who are included in the Lister-Ohlsen sample, but omitted from the present study, were those of a school district used for control purposes in the original study. The only teacher ratings and student data gathered in this school district were those used in an investigation of the inter-judge reliabilities of the various teacher ratings obtained.

The students were enrolled in four rural public school systems located near the Champaign-Urbana campus of the University of Illinois, and represented four grade levels: fifth, seventh, ninth, and eleventh.

The 48 teachers from whom ratings were obtained were teachers of the students described above. In grades five and seven they were teachers of self-contained classrooms. In grades nine and eleven the teachers all taught academic subjects (e.g. mathematics, English, social studies, foreign languages and various sciences).

Data

The data for the present study were of three types: first, the results of the Henmon-Nelson Test of Mental Ability, administered as a part of the Lister-Ohlsen study; second, ratings and estimates obtained by the present investigator from teachers; third, students' self-estimates of their intellectual ability secured by Lister and Ohlsen.

The Henmon-Nelson was originally selected by Lister and Ohlsen because it combined speed of administration with high validity and reliability in providing a global measure of intelligence. Lefever (1959) has reported validity and reliability information relating to the Henmon-Nelson. He reports correlation coefficients between Henmon-Nelson IQ's and total achievement battery scores ranging from .64 to .85, with a median value of .79. Coefficients indicating the relationship between Henmon-Nelson IQ's and average teacher grades were reported as having a median value of .60 and a range of from .09 to .74. For several grade levels examined, correlation between IQ's and various achievement subtests ranged from .65 to .76. Also, for several grade levels the median correlation coefficients between Henmon-Nelson IQ's and teachers grades were found to range from .53 to .71. As a global measure of general intelligence the Henmon-Nelson was especially well-suited to the purposes of the present study in which it was used as the criterion against which teacher estimates of general student intellectual ability were compared.

Teacher ratings were obtained on rating sheets (see rating form at the end of this chapter) which asked teachers to rate their students in the following areas: personal appearance, classroom conduct, social and economic background, intellectual ability and academic achievement. The scale on which teachers rated the intellectual ability of their students was purposely located with the other scales in order to prevent any unusual attention being called to it. With reference to the rating of intelligence, teachers were asked to make their estimates along a rating scale provided with descriptive statements and rough percentile equivalents, rather than estimating an IQ score which the pupil might earn on an intelligence test. This was deemed especially necessary due to the fact that in most of the schools included in the study the teachers had normal access to existing test data from the schools' own testing programs, gathered prior to the present study. However, more availability of test results does not necessarily mean that such results had been seen or used by the teachers. Even if teachers had, however, scrutinized prior, existing test data on their students, the aims of the present study are not affected. In such cases what would have been examined were factors associated with teacher rating inaccuracy which persisted in spite of the availability of test information. It could even be argued, perhaps, that the examination of teacher ratings under such conditions bears more relevance to everyday educational practice since teachers typically do have access to test data relating to their students. The present investigators believe that requiring teachers to express their estimates of students' intellectual ability in rough percentile rank equivalents assured, as much as is possible, that the estimates made by teachers were estimates and not just the parroting back of test information.

The ratings made by teachers in the other four areas were also made on a scale provided with descriptive statements and percentile rank bands. Again, it is believed that asking teachers to express their ratings of students in terms of a percentile scale provided them with a conceptual framework with which they were familiar and which, because it can be understood in terms of simple numerical or quantitative concepts, lends itself to the description of individual differences.

As has been previously mentioned, a pilot investigation was carried out to determine the reliabilities of the various ratings obtained. This investigation was carried out using six teachers of 11th grade students. Three groups of 30 students each were selected. For each group, ratings were obtained from two teachers. The small number of teachers involved in this pilot study was necessitated by the fact that in small school districts the number of different teachers teaching academic subjects at a particular grade level is restricted; therefore, it was difficult to find more than two or three different teachers of academic subjects who came into contact with all of any of the groups of students. The following median inter-judge reliability coefficients were obtained for the various areas of the teacher rating scale: personal appearance .463, classroom behavior .112, social and economic background .792, intellectual ability .610, and academic achievement .781.

All of the original raw data were converted to normally-distributed *T*-scores. In the case of Henmon-Nelson scores this was done by using the percentile rank equivalents of students' scores as the basis for conversion into *T*-scores. Teachers' ratings were converted into *T*-scores by finding the *T*-score equivalent of the mid-percentile rank value of the division of the rating scale in which the rating was made. This, of course, assumed that the teachers, in making their ratings, were doing so with reference to a normal distribution as the directions for the rating scales asked them to do. Admittedly, this assumption is open to some question; however, it would, perhaps, be just as questionable to assume that the teachers were making their ratings in accord with some other distribution, for example the actual distribution of the variable in *their* classroom. Furthermore, with such variables as personal appearance or classroom behavior, no objective, quantitative criteria are available to suggest either the distribution or the variable in a particular classroom or in the total population; therefore, the assumption of a normal distribution seems no more arbitrary than would be the assumption of any other distribution.

Students' self-estimates of intellectual ability were made along a rating scale which, like that used by their teachers was divided into percentile bands. Again, the mid-percentile rank value of the division of the rating scale in which the student placed himself was used as the basis for converting his self-estimate into a *T*-score.

Results

Hypothesis 1

> There will be a significant, positive relationship between discrepancy scores representing the degree and direction (i.e. under- or over-estimation) of discrepancy between teacher estimates of students' intellectual ability and the students' actual performance on the Henmon-Nelson and teacher ratings in each of the following areas: 1) personal appearance, 2) classroom conduct, 3) social and economic background, and 4) academic achievement.

Data

The data used in the analysis for this hypothesis were of two types: TEACHER ratings of their students with reference to personal appearance, classroom conduct, social and economic background and academic achievement and discrepancy scores representing the degree and direction of teacher inaccuracy in estimating the same students' intellectual ability.

Teacher ratings of students in the four areas were obtained on rating scales which were divided into decile bands and provided with descriptive statements. Teacher ratings in each of the four areas were transformed into _T_-scores.

Discrepancy scores were obtained by subtracting the _T_-score equivalent of a student's Henmon-Nelson performance from the T-score equivalent of his teacher's estimate of his intellectual ability, and adding the resulting figure to a constant of 30. Thus, a discrepancy score larger than 30 was indicative of teacher overestimation of a student's intellectual ability, while a discrepancy score smaller than 30 indicated underestimation.

Statistical Procedures

Pearsonian product-moment correlation was used in the analysis for this hypothesis. Teacher ratings in each of the four areas were correlated with the discrepancy score, resulting in four coefficients. This procedure was followed for the total sample and also separately for students at each grade level. The resulting coefficients were then examined to determine if they differed significantly from .00.

Results

Results of the analysis for this hypothesis are presented in Table 1.

A number of these coefficients are large enough to differ significantly from .00. However, with the possible exception of academic achievement, coefficients in the four areas were neither sufficiently consistent nor of a high enough order to unequivocally support the hypothesis.

Discussion of the Results

This hypothesis was set up to test the expectation that as discrepancy scores increased above the constant value of 30 (i.e. as teachers overestimated students' intellectual ability) there would be a significant tendency for teachers' ratings in each of the four areas investigated to also increase, and, similarly, as discrepancy scores decreased below the constant value of 30 (i.e. as teachers underestimated students' intellectual ability) there would be a significant tendency for teachers' ratings in each of the four areas to decrease. This expectation was based, generally, upon the idea that teachers react to students in terms of a general perceptaul "set" and that a positive or negative reaction to a student with reference to a particular characteristic will spread over and be reflected in her perception of the student with regard to other characteristics. Thus, for example, if a student is well-dressed and comes from a "good" family, this point-of-view would expect him to have a

TABLE 1

RELATIONSHIP BETWEEN TEACHERS' RATINGS OF
STUDENTS IN FOUR DESCRIPTIVE AREAS, AND DISCREPANCY SCORES

		Correlation, Henmon-Nelson vs.			
Group	N	Personal Appearance	Classroom Conduct	Social & Economic Background	Academic Achievement
Grade 5	160	.153	.126	.178*	.279**
Grade 7	141	-.031	.008	.140	.080
Grade 9	127	.084	-.037	.118	.196*
Grade 11	134	.392**	.283**	.099	.376**
Total N	562	.116**	.058	.125**	.268**

*P < .05
**P < .01

tendency to perceive him positively with regard to other related or unrelated characteristics, for instance: honesty.

As was previously pointed out, the results of the analysis for this hypothesis do not clearly support this position, at least with regard to over- and/or underestimation of students' intellectual ability by teachers. Only with reference to the area of academic achievement was there noted anything approaching a consistent pattern of significant, positive correlation between the two variables. Some basis for this latter finding may lie in the idea that a teacher, in making his judgment of a given student's intellectual ability, would depend heavily upon the student's academic achievement in the classroom as an indicator of the level of this intelligence. Therefore, as students' classroom achievement goes up, even though as a function of such factors as motivation rather than basic ability, the teacher would tend to see this as indicative of an increasing level of intellectual ability. Conversely, as students' academic performance declines, even though due to such an influence as lack of interest, the teacher would be apt, in many cases, to see this as reflecting lack of general intellectual ability. Obviously, this position is a gross generalization since there are many teachers who are sophisticated enough to be able to separate the various sorts of influences which may interact to produce a given student's academic achievement. However, the consistency of a significant coefficient between the two variables (i.e. discrepancy scores and ratings of academic achievement) suggests the operation of some tendency for teachers to overestimate the intellectual ability of students as their ratings of the students' academic achievement increase--and _vice versa_. The size of the coefficients does not, however, point to _its influence_ as having been strong in the ratings of _all_ teachers for _all_ students.

In accordance with the above line of speculation, the finding of less significant relationship between discrepancy scores and teacher ratings in each of the other three areas might indicate that such factors as personal appearance or classroom behavior (being essentially non-intellective characteristics) played a less important part in affecting a teacher's judgment of a student's intellectual ability.

Further clarification of the findings may be obtained by inspecting the intercorrelations, for the entire sample, among the five areas in which teacher ratings were obtained.

TABLE 2

INTERCORRELATION MATRIX FOR THE FIVE AREAS IN WHICH TEACHER RATINGS WERE OBTAINED

	Area 1	Area 2	Area 3	Area 4	Area 5
Area 1 (Personal Appearance)	--	--	--	--	--
Area 2 (Classroom Conduct)	.677	--	--	--	--
Area 3 (Soc. & Econ. Background)	.629	.504	--	--	--
Area 4 (Intellectual Ability)	.589	.563	.593	--	--
Area 5 (Academic Achievement)	.596	.621	.600	.906	--

Generally, it may be observed that these coefficients are consistently of a much higher order than those found in the analysis for the original hypothesis, between teachers' ratings in the four areas and discrepancy scores. These intercorrelation coefficients would be consistent with the idea that teachers react to students in terms of a general "set" which is reflected in their judgments of the students with regard to many different characteristics. However, they would also be consistent with the notion that the main factor is not a "set" which glosses over and distorts the degree to which students actually possess a given characteristic, but that teachers are, in reality, perceiving and reporting accurately, and that there is actually a high relationship among these characteristics. To discover such cause and effect relationships is, or course, beyond the scope of the present investigation; however, whatever the basis of the observed relationships, it seems that with the possible exception of the area of academic achievement, it does not extend to the over- and/or underestimation of student intellectual ability as represented by discrepancy scores. At least, such a relationship does not seem to exist across the entire sample, as seems to be the case with the other variables. Perhaps this is a function of the fact that in general teachers were quite accurate in estimating the intellectual ability of their students. One could speculate that it is only with those students whose discrepancy scores are extremely large that the expected relationships between discrepancy scores and teacher ratings for the various areas become operative.

Hypothesis 2, examining as it does students rated at the extremes with reference to the four areas, is designed to provide information to support or reject this latter line of speculation.

Hypothesis 2

> There will be significant, systematic differences between the mean discrepancy scores (scores which represent the degree and direction of discrepancy between a teacher's estimate of a student's intellectual ability and the student's actual performance on the Henmon-Nelson) of students rated high (i.e. $> +1$ s.d.) and those rated low (i.e. < -1 s.d.) in each of the four areas mentioned in Hypothesis.1. More specifically:
>
> a. the discrepancy scores of students rated high in the area of personal appearance will be significantly higher (and will indicate overestimation) than those of students rated low (whose discrepancy scores will indicate underestimation),
>
> b. the discrepancy scores of students rated high in the area of classroom conduct will be significantly higher (and will indicate over estimation) than those of students rated low (whose discrepancy scores will indicate under-estimation),
>
> c. the discrepancy scores of students rated high in the area of social and economic background will be significantly higher (and will indicate over-estimation) than those of students rated low(whose discrepancy scores will indicate under-estimation),
>
> d. the discrepancy scores of students rated high in the area of academic achievement will be significantly higher (and will indicate over-estimation) than those of students rated low (whose discrepancy scores will indicate under-estimation).

Data

Data for the analysis of this hypothesis were of two types: T-score equivalents of teachers' ratings of students in each of the four areas, and discrepancy scores representing the degree and direction of discrepancy between teachers' estimates of students' intellectual ability and the students' actual performance on the Henmon-Nelson. The derivation of these discrepancy scores has previously been described in connection with Hypothesis 1.

Statistical Procedures

The t-test of mean difference was used to test all parts of this hypothesis. For each of the areas investigated, two groups of students were drawn from the sample: the first group was made up of students whose teacher-assigned rating in that particular area was equal to a T-score of

66 (i.e. teachers had rated them in the highest rating category with respect to the area), the second group consisted of students whose rating in that area was equal to a T-score of 39 or 34 (i.e. teachers had rated them in one of the lowest two categories with respect to this area). It should be noted that the upper group was composed of students whose ratings fell in one rating category, while the lower group was composed of students whose ratings fell in two rating categories. This method of forming the group was necessary because of the fact that the distributions of teachers' ratings in all four areas were skewed in a negative direction; that is, most of the ratings fell toward the high end of the scales. Therefore, to get enough cases in the lower group to allow statistical procedures to be used, it was necessary to include those students whose ratings fell in the lower two categories.

After the two groups to be compared had been selected for each area, Bartlett's test for homogeneity of variance was applied to the variances of the two groups to determine if it were reasonable to consider the two populations equally variable. In all cases the results of Bartlett's test indicated that the two groups could not be considered equally variable. This necessitated that the t-test of mean difference which was ultimately used to test for significance of differences between the mean discrepancy scores of the two groups in each area, use separate rather than pooled variances.

Also, for areas 1, 2 and 3 the relatively small size of the lower group made it necessary to apply the method suggested by Cochran and Cox to determine the critical t value for testing the results of the t-test analysis. Only for Area 1 was it necessary to actually calculate the critical t value, since in the other cases the resulting t-values either exceeded or fell short of the theoretical limits for the critical t-values.

Results

The results of the analysis for this hypothesis are presented in Tables 3 through 6. Each sub-section of the main hypothesis will be considered separately.

2a. Difference between the mean discrepancy scores of students rated high (i.e. T-score equal to 66) and those rated low (i.e. T-score equal to 39 or 34 in the area of personal appearance.

TABLE 3

t-TEST OF MEAN DIFFERENCE BETWEEN THE DISCREPANCY SCORES OF STUDENTS RATED HIGH AND THOSE RATED LOW IN THE AREA OF PERSONAL APPEARANCE

Group	N	$\bar{X}$	s.d.	t	$t_{.05}$
High	84	31.87	7.5	2.25	2.19
Low	12	23.50	13.09		

Table 3 presents the results of the analysis which was carried out to determine whether or not these two groups differed significantly with reference to mean discrepancy score.

These results support the hypothesis since the obtained *t* value exceeds the critical *t* value obtained by application of the Cochran and Cox formula. It should also be noted that the mean for the low group is indicative of underestimation, while the mean for the high group is only somewhat higher than the theoretically "accurate" score of 30.00.

2b. Difference between the mean discrepancy scores of students rated high (i.e. T-score equal to 66) and those rated low (i.e. T-score equal to 39 or 34) in the area of classroom behavior.

Table 4 presents the results of the analysis which was carried out to determine if these two groups of students differed significantly in their mean discrepancy scores.

TABLE 4

t-TEST OF MEAN DIFFERENCE BETWEEN THE DISCREPANCY SCORES OF STUDENTS RATED HIGH AND THOSE RATED LOW IN THE AREA OF CLASSROOM BEHAVIOR

Group	N	$\bar{X}$	s.d.	*t*	*p*
High	130	29.72	6.5	1.66	NS
Low	25	26.20	10.2		

Since the results of this analysis do not indicate a significant difference between the mean discrepancy scores of these two groups of students, the hypothesis is not supported.

2c. Difference between the mean discrepancy scores of students rated high (i.e. T-score equal to 66) and those rated low (i.e. T-score equal to 39 or 34) in the area of social and economic background.

TABLE 5

t-TEST OF MEAN DIFFERENCES BETWEEN THE DISCREPANCY SCORES OF STUDENTS RATED HIGH AND THOSE RATED LOW IN THE AREA OF SOCIAL AND ECONOMIC BACKGROUND

Group	N	$\bar{X}$	s.d.	*t*	*p*
High	37	29.86	5.3	2.49	$< .05$
Low	24	23.67	12.3		

Table 5 presents the results of the analysis which was carried out to determine if these two groups of students differed significantly with respect to their mean discrepancy scores.

The hypothesis is supported since the results of the analysis indicate a significant difference between the two groups of students with reference to their mean discrepancy scores.

An inspection of the means also indicates that while teachers were quite accurate in estimating the intellectual ability of the group rated high in the areas of social and economic background, there was a tendency for the teachers to underestimate the intellectual ability of students rated at the low end of the continuum on this area.

2d. Difference between the mean discrepancy scores of students rated high (i.e. T-score equal to 66) and those rated low (i.e. T-score equal to 39 or 34) in the area of academic achievement.

Table 6 presents the results of the analysis carried out to test for the existence of significant difference between the mean discrepancy scores of these two groups of students.

TABLE 6

t-TEST OF MEAN DIFFERENCE BETWEEN THE DISCREPANCY SCORES OF STUDENTS RATED HIGH AND THOSE RATED LOW IN THE AREA OF ACADEMIC ACHIEVEMENT

Group	N	$\bar{X}$	s.d.	t	p
High	57	32.18	5.1	4.72	$<.01$
Low	56	24.95	10.3		

These results, since they indicate the presence of significant difference between the mean discrepancy scores for the two groups of students, dictate the retention of the hypothesis.

An inspection of the two means in question indicates that the mean for the group of students rated high differs from the theoretical "no discrepancy" score of 30.00 in the direction of overestimation, while the group rated low differs in the direction of underestimation. However, the latter group is more discrepant, even taking into consideration the fact that its distribution has a greater variance than the former group.

Results

With the exception of Area 2 (classroom behavior) the results of the above analyses indicate support of the hypothesis that there would be found significant, systematic differences between the discrepancy scores of students rated high and those rated low in each of the four areas.

This hypothesis was designed to explore the idea that when a student is perceived by his teacher as being superior in comparison to most other students with reference to one or all of the rating areas, there would be a tendency for the teacher to overestimate his intellectual ability; similarly, when a student is perceived as inferior with respect to these areas there would be a tendency for the teacher to underestimate his intellectual ability. This means that the hypothesis demanded more than just the finding of significant difference between the two mean discrepancy scores: it also required that the mean of the high group be in excess of 30.00 (the constant to which was added the result of subtracting the T-score equivalent of the student's Henmon-Nelson performance from the T-score equivalent of his teacher's estimate of his intellectual ability) and that the mean of the low group in each area be less than 30.00.

When the results are examined to determine if the high group in each area was, indeed, overestimated and the low group underestimated, the expected pattern seemed to be borne out only for the group of students rated at the low end of the continuum with respect to the various characteristics; that is, for this group of students the mean discrepancy scores were indicative of underestimation. However, the mean discrepancy scores of the students rated high by their teachers in each area were much closer to 30.00, and not nearly as consistent or strong a pattern of overestimation was noticed. Stating this finding generally, it can be said that while students who were perceived as being extremely inferior with reference to each of the rating areas tended to have their intellectual ability underestimated by their teachers, there was no strong indication of the opposite pattern for students who were perceived by their teachers as being extremely high in each of the rating areas.

Attempting to tie these findings in with the idea that teachers perceive and respond to students on the basis of a perceptual set which cuts across many of the students' characteristics and areas of functioning, it would generally seem that only the operation of negative set is indicated. That is, with three of the areas investigated there was a definite tendency for teachers to underestimate the intellectual ability of students whom they perceived as being markedly inferior with respect to the characteristics being rated; however, for none of the areas was there a similar, marked tendency for teachers to overestimate the ability of students perceived as superior with reference to the same areas.

If we project these findings into the classroom situation, it would seem that the student who is seen by his teachers as being markedly inferior with respect to his personal appearance, social and economic background, or academic achievement is the type of student whom they are apt to perceive as having less ability than is indicated by group tests of intelligence. One might anticipate that students who would be more apt to meet one or all of these criteria would be students coming from a very low socioeconomic background. Furthermore, if we may suppose that a teacher's general judgment of a student's intellectual ability plays an important part in determining the level and quality of academic performance of which he will see the student as being capable, and if these expectations, in turn, are subtly or more directly communicated to the student, then we may have a situation in which the lower class student, coming from a home which generally gives him little encouragement to do well academically in the school situation, is met by teachers who reinforce, or at least contribute to, lower levels of aspiration than his ability would warrant.

It is interesting to speculate why significant and systematic differences between the two groups of students were not found in connection with the area of classroom behavior. One possible explanation for this may lie in there being a variety of patterns of the deviant and conforming behavior which fall at the opposite extremes of the classroom behavior continuum. For example, the quiet, passive, withdrawing behavior of the poor student who has given up, as well as the conscientious, attentive, conforming behavior of the good student may both be considered to be "good" classroom behavior by the teacher. On the other hand, both the hostile, antisocial student who is unable to meet the intellectual demands of schoolwork, and the bright, aggressive student who challenges his teachers and school procedures, may equally be seen as "bad actors" by teachers. If poor and good classroom behavior are not unitary characteristics, then there may be confounding teacher rating patterns operating in the two groups examined, precluding the possibility of finding the type of results which were obtained in connection with the other areas.

The above position receives some support from the finding, reported earlier in describing the teacher rating form, that the interjudge reliability coefficients for classroom behavior were lower (median value of .112) than for any other rating area. There was, then, least agreement among the teachers with reference to their ratings of the classroom behavior of the same students. Perhaps this indicates that they were operating under different criteria of what constitutes good and poor behavior in the classroom.

Conclusions

The following conclusions may be drawn from the results of the analyses:

1. Across the entire sample, there was generally no definite tendency for teachers to overestimate the intellectual ability of students as their ratings of these students in the several areas went up, and vice versa.

2. With the exception of classroom behavior, there are significant differences between the mean accuracy scores of students rated high in each of the descriptive areas and those rated low. While teachers generally underestimated the intellectual ability of students rated "Low" in each of the areas, there was no corresponding tendency for them to overestimate the intellectual ability of those students rated high.

BIBLIOGRAPHY

Ackerson, L., and Estabrooks, G. H., "On the Correlation of Intelligence Test Scores with Imputed Intelligence," *British Journal of Psychology*, 1928, 18:455-459.

Bromse, Gerald C., "A Study of Teachers' Ability to Accurately Estimate the Intelligence Quotients of Puerto Rican Children," *Psychological Newsletter*, 1956, 7(3):37-38.

Gage, N. L., Leavitt, G. S., and Stone, G. C., "Teachers' Understanding of their Pupils and Pupils' Ratings of Their Teachers," *Psychological Monographs*, XXI, (1955), 37.

Gibbons, C. C., "A Comparison of Kuhlmann-Anderson Test Scores and Teachers' Estimates," *School and Society*, 1938, 47:710-712.

Hanna, J. V., "Estimating Intelligence by Interview," *Educational and Psychological Measurement*, 1950, 10 (3): 420-430.

Hubtard, R. E., and Flesher, W. R., "Intelligent Teachers and Intelligence Tests, Do They Agree?" *Educational Research Bulletin*, May, 1953, 32:113-122.

Jordan, A. M., "The Validation of Intelligence Tests," *Journal of Educational Psychology*, 1923, 14:348-366.

Lefever, D. Welty, in *The Fifth Mental Measurements Yearbook*, Oscar K. Buros, ed. Highland Park, The Gryphon Press, 1959, 470-472.

Lewis, W. Drayton, "Some Characteristics of Dhildren Designated as Mentally Retarded, as Problems, and as Geniuses by Teachers," *Journal of Genetic Psychology*, 1947, 70:29-51.

Lister, J. L., and Ohlsen, M. M., *The Effects of Orientation to Testing on Motivation for and Outcomes of Test Interpretation*, Cooperative Research Project No. 1344, College of Education, University of Illinois, 1962.

Moore, C. C., "Note on the Accuracy of Judgment in Pupil Ability Ratings," *Journal of Educational Psychology*, 1937, 28:391-393.

Olander, H. T., and Walker, B. S., "Can Teachers Estimate IQ's?" *School and Society*, 1936, 44:744-746.

Schwarz, J., "Szacowanie Inteligencji," *Psychometrja*, 1934, 1:17-30.

Wilson, J. H., "Comparison of Certain Intelligence Tests," *British Journal of Psychology*, 1925, 15:44-63.

Wilson, J. Watson, "Correlation of Clinical Estimates with Test Scores on Mental Ability and Personality Tests," *Journal of Clinical Psychology*, 1954, 10:97-99.

TEACHER'S MANUAL AND RATING SHEETS

______(Teacher's Name)__________,

In connection with the research project being carried on in your school at the present time, you are asked to rate a group of your students with regard to the following areas:

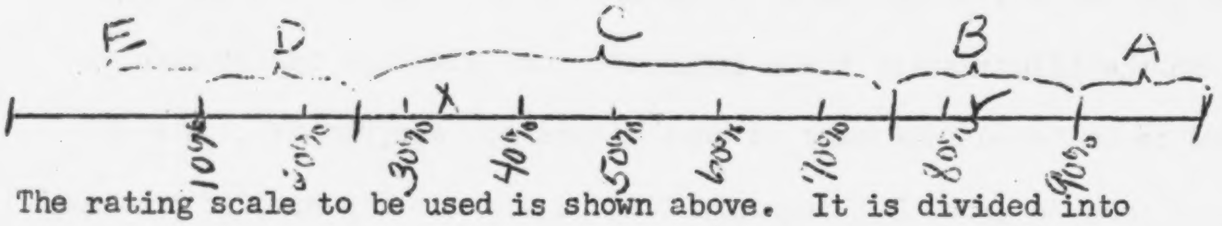
1. personal appearance
2. classroom conduct
3. social and economic background
4. academic achievement

Definitions of these terms and a description of the rating scale to be used will be found on the following pages.

Realizing how limited a commodity time is to the average teacher an attempt has been made to keep the number of rating categories at a minimum and the rating scale as uncomplicated as possible. I think you will find that the process of rating these students will take you only a short time.

The information is being gathered for research purposes and each teacher's ratings will, of course, be kept strictly confidential. These data will be made available only to those who will have the responsibility for analyzing them at a later phase of the project.

Richard E. Pearson
Graduate Research Assistant
University of Illinois



The rating scale to be used is shown above. It is divided into five broad segments: A; B; C; D; and E. Descriptively, a check in each of these segments will mean the following:

A - with regard to the characteristic being rated, the student is judged to be much higher than most other students his age

B - with regard to the characteristic being rated, the student is judged to be somewhat higher than most other students his age

C - with regard to the characteristic being rated, the student is judged to be similar to most others his age; that is, he or she is about average

D - with regard to the characteristic being rated, the student is judged to be somewhat lower than most other students his age

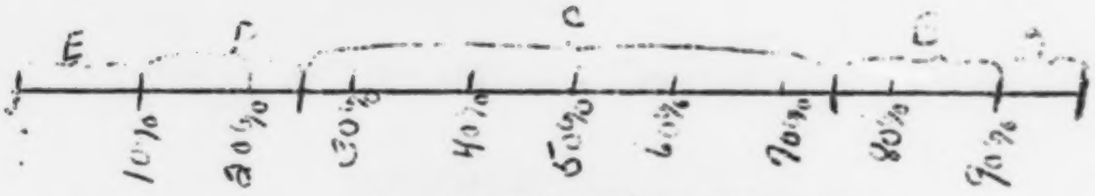
E - with regard to the characteristic being rated, the student is judged to be much lower than most other students his age

You will notice that in segments B, C, and D there are smaller segment portions. Each of these subsegments is intended to represent about 10% of the total population being considered, Note that the upper and lower boundaries of segment C each fall mid-way in one of sub-segments. While rating a student on a particular characteristic, please use the sub-segments to make sharper distinctions than would be indicated by simply putting a check in the middle of one of the broad descriptive segments. For example, the "✓" on the illustrative scale on page 2 indicates that although the student is rated as "somewhat higher than most students his age," with regard to the characteristic being considered, he is closer to group A than group C. Similarly, the "x " on the illustrative scale indicates that although the student is judged to be "about the same as most others his age,"he is closer to group D than to group B. Segments A and E are not sub-divided because in each case the whole segment represents 10% of the total population being considered.

Given below are definitions of the characteristics upon which you are asked to rate your students:

1. Personal appearance - this refers to the extent to which the student's clothes and person reflect a concern for cleanliness and neatness. This is not, however to be interpreted as a judgment of the quality or value of the student's clothing; rather, what is desired in your judgment of the extent to which the student uses what he has to present a pleasing appearance,
2. Classroom conduct - this refers to the usual behavior of the student while he is in your classroom. Such things as the student's orderliness, attentiveness and co-operative attitude (or lack of them) should be the basis of your judgment. Stated generally, this category can be described as the extent to which the student adheres to your rules of good classroom behavior,
3. Social and economic background - this refers to your judgment of the general standing in the community of the student's family with regard to such factors as: occupation and income level of the parent(s); type and area of residence; participation of family in the social and civic functions of the community,
4. Intellectual Ability - this refers to the capacity which the student has to meet such intellectual demands as those posed by school learning,
5. Academic achievement - this refers to the extent to which the student has learned the subject matter covered in your classes. Pupils will, of course, have different levels of academic achievement in different areas, so what you are asked to make is a judgment of the student's over-all achievement in your classes.

Each teacher, as a result of his or her day-to-day contact with students forms informal opinions about them. In all of these ratings what is desired is an estimate based on these informal opinions.



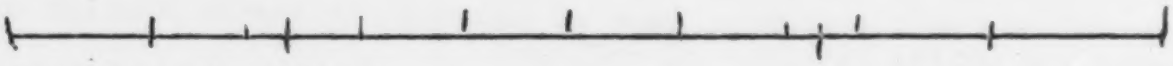
A - is much higher than most
B - is somewhat higher than most
C - is average
D - is somewhat lower than most
E - is much lower than most

Student's Name____

1. Personal appearance:

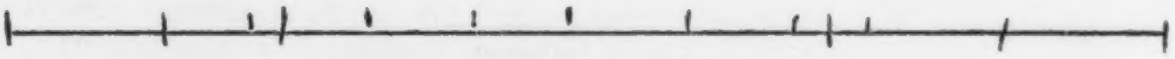
2. Classroom conduct:

3. Social and economic background:



4. Intellectual ability:

5. Academic achievement:



THE RELATIONSHIP BETWEEN TEACHERS' ESTIMATES OF STUDENTS' INTELLECTUAL ABILITY AND THE STUDENTS' SELF-ESTIMATES

Richard E. Pearson and Paul A. Wurm

Objectives

Test interpretation may be described as being primarily a matter of focusing upon an individual's self-perceptions. The counselor attempts to clarify, confirm or change the individual's self-perceptions in the light of obtained test data. The present paper has as its objective the exploration of the relationship between students' self-perceptions of their intellectual ability and an influence which may play an important part in the formation or modification of these self-perceptions -- namely, teachers' judgments of the students' intellectual ability.

The relationship between teachers' estimates of students' intellectual ability and students' own self-estimates is considered to be one of importance because in the student's environment teachers certainly represent one of the primary sources of "feed-back" concerning his intellectual ability. Directly or indirectly, many of the teacher's interactions with a student can be seen as communicating his implicit judgment of the student's intellectual ability. Because of the teacher's position of authority and the frequency of his interaction with the student, it is believed by the present investigators that these implicit judgments, internalized by the student, have an important influence upon his own perception of his intellectual ability.

The objective of exploring the relationship between teachers' estimates of students' intellectual ability and the students' self-estimates will be pursued by means of the following hypotheses, stated in the null form:

Hypothesis 1. No significant, positive product-moment correlation will be found between teachers' estimates of students' intellectual ability and the students' actual performance on the Henmon-Nelson Test of Mental Maturity.

Hypothesis 2. The product-moment correlation coefficient between teachers' estimates of students' intellectual ability and Henmon-Nelson results will not be significantly higher than that between students' own self-estimates of intellectual ability and their Henmon-Nelson performance.

Hypothesis 3. The product-moment correlation coefficient between students' self-estimates and Henmon-Nelson performance will not differ significantly from that between these self-estimates and teachers' estimates of the students' intellectual ability.

Hypothesis 4. There will be no significant relationship between the classification of students as "over-estimated", accurately estimated", and "under-estimated" made on the basis of the difference between students' Henmon-Nelson performance and teachers' estimates, and similar classification made on the basis of the difference between Henmon-Nelson performance and students' self-estimates.

Related Research

Since this paper essentially represents an area of inquiry which overlaps both of the preceding papers, the relevant research has already been reported in their reviews of the research. To summarize briefly, studies relating to students' estimates of their intellectual ability indicate that more intelligent pupils tend to estimate their test scores more accurately than do the less intelligent, (Froelich and Moser, 1954; Lister and Ohlsen, 1962) and that students who grossly over- or under-estimate their abilities appear to be less intelligent and less well adjusted than students who are more accurate in their self-estimations. (Arsenian, 1942).

Also, as has been previously mentioned, studies investigating the relationship between teachers' estimates of students' intellectual ability and the students' actual performance on a criterion measure have generally found high, positive correlation, with coefficients (both Pearsonian and rank-order) ranging from about .40 to .75. (Ackerson and Estabrooks, 1928; Wilson, 1925; Schwarz, 1934; Olander and Walker, 1936; Hubbard and Flescher, 1953; Moore 1937; Jordan, 1923.)

Procedures

Sample

The sample used in this study is the same as that of the previous study (see p 63).

Data

There were three types of data for the present study: first, students' Henmon-Nelson results, gathered as a part of the data for the Lister-Ohlsen study (1962); second, students' estimates of their intellectual ability, also gathered by Lister and Ohlsen; third, teachers' estimates of students' intellectual ability gathered by Pearson (see p.77 for an example of the rating scale upon which these rating scales were obtained).

The original raw data were converted to normally-distributed T-scores in accordance with the procedures described in the previous paper (see p. 69).

Significance Level

The .05 level of significance is accepted in evaluating the results of the analyses carried out. In cases where the results surpass this level, such is indicated in the presentation of the results.

Results

Hypothesis 1

No significant, positive product-moment correlation will be found between teachers' estimates of students' intellectual ability and the students' actual performance on the Henmon-Nelson.

Data Used in the Analysis for Hypothesis 1

The two variables which were correlated in the analysis of this hypothesis were: *T*-score equivalents of students' actual performance on the Henmon-Nelson, and *T*-score equivalents of teachers' estimates of the students' intellectual ability.

Statistical Procedures Used in the Analysis for Hypothesis 1

Pearsonian product-moment correlation was used in the analysis of this hypothesis. The resulting coefficient was then tested to determine if it differed significantly from .00.

Results of the Analysis for Hypothesis 1

Results of the analysis for this hypothesis are presented in Table 1.

TABLE 1

PRODUCT-MOMENT CORRELATION BETWEEN TEACHERS'
ESTIMATE OF STUDENTS' INTELLECTUAL ABILITY AND THE
STUDENTS' HENMON-NELSON PERFORMANCE

Group	N	r	p
Total Group	562	.682	--

These results indicate that for the total sample, there was a high degree of positive relationship between standard score equivalents of students' performance on the Henmon-Nelson and teachers' estimates of the students' intellectual ability.

Discussion of the Results

The body of research examining teachers' ability to estimate the intelligence of their students, and the fact that teacher ratings of students' intellectual and academic ability (generally in the form of assigned grades) have served as one of the chief validating criteria for intelligence tests, would lead one to expect the high, positive relationship indicated by the results of this analysis.

The results indicate that the estimates which teachers made of their students' intellectual ability were highly related to the students' performance on a group test designed to provide a global index of general intelligence. The coefficient obtained is quite consistent with those of earlier studies which used similar techniques.

Going beyond the analysis carried out for the hypothesis, when the data were examined to determine the accuracy of teachers' estimates, a mean absolute discrepancy of 5.63 T-score points were found between the teachers' estimates and actual Henmon-Nelson results.

Hypothesis 2

> The product-moment correlation coefficients between teachers' estimates of students' intellectual ability and Henmon-Nelson results will not be significantly higher than that between students' own self-estimates of intellectual ability and their Henmon-Nelson performance.

Data

There were three variables used in the analysis for Hypothesis 2: students' Henmon-Nelson performance, teacher estimates of students' intellectual ability, and students' self-estimates of their intellectual ability. Henmon-Nelson results were converted to T-scores on the basis of percentile rank equivalents of raw scores. Teachers' estimates and students of intellectual ability were also converted to T-scores by means of the percentile rank bands of the rating scales within which these estimates were made.

Statistical Procedures

Statistical procedures used in the testing of Hypothesis 2 consisted of Pearsonian product-moment correlation, and the t-test to determine the significance of difference between two correlation coefficients. The product-moment coefficient was computed to determine the relationship between students' Henmon-Nelson performance and teachers' estimates of the students' intellectual ability. The coefficient was also computed between students' Henmon-Nelson performance and their self-estimates of intellectual ability. The resulting coefficients were then tested to determine if they were significantly different.

Results

Results of the analysis for Hypothesis 2 are presented in Table 2. For the total group, teacher estimates of students' intellectual ability were correlated with Henmon-Nelson performance to a significantly higher

TABLE 2

SIGNIFICANCE OF DIFFERENCES BETWEEN PRODUCT-MOMENT
CORRELATION OF HENMON-NELSON RESULTS AND TEACHERS'
ESTIMATES OF STUDENTS' INTELLECTUAL ABILITY,
AND HENMON-NELSON RESULTS AND STUDENTS' SELF-ESTIMATES

	Correlation			
Group	Henmon-Nelson vs. Teacher Estimate	Henmon-Nelson vs. Self-Estimate	*t*	*p*
Total Sample	.682	.587	5.87	$< .01$

degree than were students' own self-estimates of intellectual ability. Therefore, the hypothesis is rejected.

Discussion of the Results

While the results indicate that teachers' estimates of students' intellectual ability bear a significantly higher degree of positive relationship to Henmon-Nelson performance than do students' own self-estimates of intellectual ability. Therefore, the hypothesis is rejected.

Discussion of the Results

While the results indicate that teachers' estimates of students' intellectual ability bear a significantly higher degree of positive relationship to Henmon-Nelson performance than do students' own self-estimates, the size of the coefficient in the latter instance is, nevertheless, indicative of a high relationship. This means that the students were able to assess their intellectual ability, in comparison to others, to a degree that is significantly greater than chance and to a degree nearly as great as their teachers.

To the extent that both teachers and students were using the same source (s) of data as the basis for their estimates, these results suggest that generally teachers' estimates are to generally be favored in making various educational decisions which rest upon an informal (i.e. no recourse is made to test data) judgments of the students' intellectual ability. However, if students and teachers were using different sources of data as a basis for their estimates (the correlation between teachers' estimates and students' estimates is .538), then perhaps a combination of these two estimates would produce more satisfactory decisions in such situations. This line of speculation receives some support when one notes that the multiple correlation coefficient between Henmon-Nelson results and a combination of teachers' estimates and students' estimates is .78. This coefficient is larger than that between either Henmon-Nelson results and teachers' estimates (.682) or Henmon-Nelson results and students' self-estimates (.587). Hypothesis 3 is designed to pursue this line of investigation.

Supplementary analysis was carried out to determine if the findings with reference to the relationship between teachers' estimates and self-estimates, and Henmon-Nelson results, were also reflected in terms of differences in accuracy between teachers' estimates and self-estimates. Table 3 presents the results of the t-test carried out to determine if there were significant difference in accuracy between the two estimates. The two discrepancy scores upon which the analysis was carried out were obtained by computing, for each student, the absolute discrepancy between his teacher's estimate and the students' Henmon-Nelson performance, and that between his self-estimate and the Henmon-Nelson results.

TABLE 3

t-TEST OF MEAN DIFFERENCE BETWEEN ABSOLUTE DISCREPANCY OF TEACHERS' ESTIMATES OF STUDENTS' INTELLECTUAL ABILITY AND THAT OF STUDENTS' SELF-ESTIMATES

Estimate	$\bar{X}$	s.d.	t	p
Teachers'	5.63	4.60	-3.16	<.01
Students'	6.55	5.13		

These results generally correspond to those of the correlational analysis carried out to test the hypothesis; that is, just as teachers' estimates were found to bear a higher degree of relationship to the criterion than did students' self-estimates, so were teachers' estimates found to be more accurate (i.e. less discrepant).

Hypothesis 3

> The degree of relationship between students' self-estimates and Henmon-Nelson performance will not differ significantly from the degree of relationship between these self-estimates and teachers' estimates of the students' intellectual ability.

Data

The data used to test this hypothesis are identical to those described in Hypothesis 2.

Statistical Procedures

The statistical procedures used to test this hypothesis consisted of Pearsonian product-moment correlation and the t-test to determine the significance of difference between two correlation coefficients.

Results

The results of the analysis for this hypothesis are presented in Table 4.

TABLE 4

SIGNIFICANCE OF DIFFERENCE BETWEEN PRODUCT-MOMENT CORRELATION; OF SELF-ESTIMATE AND HENMON-NELSON RESULTS AND SELF-ESTIMATES AND TEACHERS' ESTIMATES

Correlation			
Self-Estimate vs. Henmon-Nelson	Self-Estimate vs. Teachers' Estimate	t	p
.587	.538	1.846	$<.05$

These results indicate that students' self-estimates are significantly more highly related to Henmon-Nelson results than they are to teachers' estimates, therefore, the hypothesis is rejected.

Discussion of the Results

The analysis for this hypothesis was undertaken with the expectation that if significant difference existed between the two coefficients, it would be in the direction of students' self-estimates being more highly related to teachers' estimates than to Henmon-Nelson results. This expectation was based upon the idea that teachers serve as one of the primary sources of "feed-back" for students with regard to information concerning their intellectual ability. From this point-of-view it would seem reasonable to expect students to rank themselves, with respect to other students, in a manner similar to that done by their teachers, even when the teacher rankings depart from those made on the basis of actual intelligence scores. Clearly, the results obtained do not support this idea. However, it should be taken into account that while in both relationships (i.e. between self-estimates and Henmon-Nelson results, and between self-estimates and teachers' estimates) the reliability of the self-estimates remains constant, the reliabilities of second variable (i.e. Henmon-Nelson performance and teachers' estimates) are not of equal magnitude. It could be that the significantly lower correlation obtained between self-estimates and teachers' estimates is an artifact due to the fact that teachers' estimates were undoubtedly less reliable than is true of the Hermon-Nelson.

Actually, test-retest reliability information was not obtained for teacher estimates, however, assuming it to be at least .60, which would be consistent with the reliability usually obtained for these kinds of ratings, correction of the obtained coefficients for attenuation suggests that the relationship between teachers' estimates and self-estimates is approximately equal to .89, whereas the relationship between self-estimates and Henmon-Nelson results is approximately .93. While it can be seen that correction for attenuation thus reduces the difference between the two coefficients somewhat, that between self-estimates and Henmon-Nelson results still remains significantly larger (t equal to 4.8, 559 d.f.) than that between self-estimates and teachers' estimates; therefore, even when correction is made for attenuation the hypothesis is still rejected.

The absence of precise information concerning test-retest reliability necessitates these letter findings be considered tentative. A future study in which complete information were gathered would be useful.

Hypothesis 4

> There will be no significant relationship between classification of students as "overestimated", "accurately estimated", and "underestimated" made on the basis of the difference between Henmon-Nelson results and teachers' estimates, and a similar classification made on the basis of the difference between Henmon-Nelson results and students' self-estimates.

Data

The data used in the analysis of this hypothesis were the classification of teachers' estimates of students' intellectual ability as "overestimated", "accurately estimated", or "underestimated", and a similar classification of students' self-estimates. Classification was made on the basis of a derived score which represented the degree and direction of the discrepancy between Henmon-Nelson results and the estimates. These discrepancy scores were obtained by subtracting the *T*-score equivalent of a student's Henmon-Nelson performance from the *T*-score equivalent of his teacher's and his own estimate of his intellectual ability and adding the resulting figures to a constant of 30. This procedure resulted in two discrepancy scores for each student, one representing the discrepancy between test results and his self-estimate, the other representing the discrepancy between test results and his teacher's estimate.

Discrepancy scores (both teachers' and students') falling at or beyond ⌿ 1 s.d. were classified as "overestimated". Discrepancy scores falling at or beyond -1 s.d. were classified as "underestimated". Finally, those discrepancy scores falling within ⌿ 1 s.d. were classified as "accurate". Thus each student was classified twice, once on the basis of his teachers' estimates and once on the basis of his own self-estimates. It was upon the distributions of these two classifications that the analysis of the hypothesis was made.

Statistical Procedures

χ^2 was used to test Hypothesis 4. Students were assigned to the appropriate cells of a contingency table on the basis of the classifications described above, yielding a 3 x 3 table for analysis.

Results

Results of the test for Hypothesis 3, presented in Table 5, indicate a highly significant relationship between the two systems of classification;

TABLE 5

RELATIONSHIP BETWEEN CLASSIFICATION OF STUDENTS' SELF-ESTIMATES AND THAT OF TEACHERS' ESTIMATES

Teacher Estimates	Self-Estimates			Total
	Over	Accurate	Under	
Over	39	65	5	109
Accurate	45	264	65	374
Under	5	44	30	79
Total	89	373	100	562

$\chi^2 = 73.3$ d.f. = 4 $p < .001$

therefore, the hypothesis is rejected.

Discussion of the Results

The purpose of this hypothesis was to expand upon the information obtained in connection with Hypothesis 3. The results of Hypothesis 3 indicate that there was a strong positive relationship between teachers' estimates of students' intellectual ability and students' self-estimates. An examination of the contingency table contained in Table 5 indicates the character of this relationship: namely, students who overestimate their own intellectual ability also tend to have their intellectual ability overestimated by their teachers; similarly, students who underestimate their intellectual ability tend to have their intellectual ability underestimated by teachers; finally, students whose self-estimates are accurate tend to have their intellectual ability accurately estimated by their teachers. It is not possible, on the basis of these findings or those of the previous hypothesis, to trace any cause-and-effect relationships among these patterns. However, one may speculate that the clear-cut tendency for association demonstrated in Table 5 was the result of the operation of one or a combination of the following influences: 1) teachers' attitudes toward students, with reference to the students' intellectual ability, serving as a model for students' self-estimates; 2) students' attitudes toward themselves, with reference to their intellectual ability, serving as a model for teachers' estimates; 3) use, by both teachers and students, of the same external sources of information (e.g. a student's intellectual repute among his peers) as a basis for their estimates. It would seem that further research in which these various influences were controlled could be carried out in order to trace the possibility of their effecting the observed relationship.

Conclusions

The hypotheses included in the present study were designed to explore the relationships among teachers' estimates of students' intellectual ability, students' self-estimates, and the students' actual Henmon-Nelson performance. Within the limitations of the study, the following conclusions may be drawn concerning the variables investigated:

1. Estimates of students' intellectual ability made both by teachers and the students themselves bear a high degree of positive relationship to the students' actual Henmon-Nelson performance,
2. Teachers' estimates are, however, significantly more highly related to Henmon-Nelson results than students' self-estimates,
3. Students' self-estimates are significantly more highly related to Henmon-Nelson scores than to teachers' estimates of the students' intellectual ability,
4. There is a high degree of association between students' self-estimates and teachers' estimates of students' intellectual ability; specifically, students who over-estimated their intellectual ability tended to be overestimated by their teachers, students who were accurate in their self-estimates tended to also be accurately estimated by their teachers, and students who underestimated themselves tended to have their intellectual ability underestimated by their teachers.

BIBLIOGRAPHY

Ackerson, L., and Estabrooks, G. H., "On the Correlation of Intelligence Test Scores with Imputed Intelligence," British Journal of Psychology, 1928, 18:455-459.

Arsenian, Seth On estimates and objective measurement. Journal of Educational Psychology, 1942, 33, 291-302.

Froelich, C. I. and Moser, W. E. Do counselees remember test scores? Journal of Counseling Psychology, 1954, 1, 145-152.

Hubbard, R. E., and Flesher, W. R., "Intelligent Teachers and Intelligence Tests, Do They Agree?" Educational Research Bulletin, May, 1953, 32:133-122.

Jordan, A. M., "The Validation of Intelligence Tests," Journal of Educational Psychology, 1923, 14:348-366.

Moore, C. C., "Note on the Accuracy of Judgment in Pupil Ability Ratings," Journal of Educational Psychology, 1937, 28:391-393.

Olander, H. T., and Walker, B. S., "Can Teachers Estimate IQ's?" School and Society, 1936, 44:744-746.

Schwarz, J., "Szacowanie Inteligencji," Psychometrja, 1934, 1:17-30.

Wilson, J. H., "Comparison of Certain Intelligence Tests," British Journal of Psychology, 1925, 15:44-63.